UNIVERSITY COLLEGE LONDON

Rainfall variability in southern Africa: drivers, climate change impacts and implications for agriculture

by

Chiara Ambrosino

A thesis submitted in partial fulfillment for the
degree of Doctor of Philosophy

in the
Faculty of Social & Historical Sciences
Department of Geography

March 2011

Declaration of Authorship

I declare that this dissertation represents my own work, except where due acknowledgement is made.
Part of the work presented in the thesis has been published:

Ambrosino, C. and Chandler, R.E. and Todd, M.C. Southern African Monthly Rainfall variability: an analysis based on generalized linear models. *J. Clim.* (2011) (On-line Early Release, doi: 10.1175/2010JCLI3924.1).

Signed: _____

Date: _____

"By means of water we give life to everything"

The Koran, Book of The Prophets 21:30

UNIVERSITY COLLEGE LONDON

Abstract

Faculty of Social & Historical Sciences
Department of Geography

Doctor of Philosophy

by Chiara Ambrosino

Rainfall variability in southern Africa: drivers, climate change impacts and implications for agriculture

Southern Africa is characterised by a high degree of rainfall variability affecting agriculture among other sectors. The focus of this study is to investigate such variability and to identify stable relationships with its potential drivers in the climate system. These relationships are used as the basis for the statistical downscaling of climate model (GCM) outputs. From the simulated rainfall, indices representative of growing season characteristics are computed with the final purpose of studying the implications on maize cropping under a future climate change scenario.

The analysis uses generalized linear models (GLMs), which allow the investigation of the relationships between different components of the climate system (geographical and climatic drivers) simultaneously. Initially, the effects of various climate indicators upon monthly regional (for all southern Africa) precipitation occurrences and amounts are characterised. Six climate factors are found to drive part of the rainfall variability in the region and their modelled effect upon rainfall occurrences and amounts agrees broadly with previous studies. Among the retained indices, relative humidity and El-Niño accounted for the highest degree of explained variability. The location and intensity of the jet stream is also found to have a statistically significant and physically meaningful effect upon rainfall variability.

Although effective for the analysis of monthly regional precipitation, and used to investigate future regional projections, the models do not perform adequately at more local spatial scales such as station locations or few km grids. The same methodology is, therefore, applied to characterise daily precipitation variability at multiple locations within a smaller region. The small scale statistical models capture adequately the seasonal and annual rainfall structure in the area. Indeed, the observations can not be distinguished from the simulated time series. However, the simulated rainfall values tend to be slightly too high throughout the seasons, possibly due to the spatial correlation structure not

completely appropriate for such a complex region.

From the simulated rainfall sequences, seven growing season indices (including the onset and length of the growing season, proportion of rainy days and total precipitation during the growing season) are derived and their projected change investigated under a climate change scenario. There is little consensus between the 18 selected GCMs, regarding changes in growing season indices between two investigated periods in the 20^{th} and 21^{st} centuries. For the next couple of decades the dominant source of variation in the indices appears to be the natural rainfall variability. Such information should therefore be taken into account when planning adaptation and mitigation strategies.

The research presented here emerges as the first comprehensive assessment of different climatic factors linked to southern Africa rainfall variability as well as the first attempt to evaluate the GLMs suitability for the generation of rainfall sequences for agricultural impact studies.

Acknowledgements

First and foremost I would like to express my gratitude to my supervisors. I am deeply grateful to Dr. Richard E. Chandler, for his patience and guidance throughout the work, and his continuous effort to explain complex things in a clear and simple way. I could not have imagined a better support and supervision in these years. I gratefully thank Prof. Martin C. Todd, for his advice and contribution, which he kept on even when new professional opportunities led him to leave the department.

I acknowledge the modelling groups, the Program for Climate Model Diagnosis and Intercomparison (PCMDI) and the WCRP's Working Group on Coupled Modelling (WGCM) for their roles in making available the WCRP CMIP3 multi-model dataset. I am thankful to Dr. Bryson Bates from CSIRO Marine and Atmospheric Research, and Dr. Mark Tadross from the University of Cape Town, Climate Systems Analysis Group, for the technical discussions and advice.

I gratefully acknowledge the UCL Environment Institute that made my Ph.D. work possible and the support I received from the Department's staff.

I thank all the friends who shared with me doctoral studies and life, in particular: Caro, Jo, Shams, Shonima, Hui, Tash, Mat, Aneta, Toby, and Gil. Thanks to my very best friends, Davide, Daniela and Fra, with them the distance that divides us never felt more than few centimetres on a map. A wise man once said: 'friendship doubles the joy and halves the sufferance'.

A special thank to my parents. They thought me to never give up and continuously challenge myself. To my twin sister, Elena, the world would not feel complete without her, it would be probably a simpler place to live in, but I would miss so many wonderful things. Thank you Ni.

Finally, to Giovanni, for the encouragement and support he gave me, for never doubting and, at times, believing more than me, I would have been lost without him.

Gio, this is for you.

Contents

List of Figures

List of Tables

Abbreviations

AMO	**A**tlantic **M**ultidecadal **O**scillation
AR4	**A**ssessment **R**eport number **4**
AGCM	**A**tmosphere **G**eneral **C**irculation **M**odel
AOGCM	**A**tmosphere-**O**cean **G**eneral **C**irculation **M**odel
BMI	**B**randon **M**arion **I**ndex
CCA	**C**anonical **C**orrelation **A**nalysis
CMIP3	**C**oupled **M**odel **I**ntercomparison **P**roject phase **3**
CRUTS3	**C**limate **R**esearch **U**nit **TS3**.0 dataset
DJF	**D**ecember **J**anuary **F**ebruary
EOF	**E**mpirical **O**rthogonal **F**unction
GCM	**G**eneral **C**irculation **M**odel
GHCN	**G**lobal **H**istorical **C**limatology **N**etwork
GHG	**G**reen **H**ouse **G**ass
GLIMCLIM	**G**eneralized **LI**near **M**odelling for daily **CLIM**ate time series
GLM	**G**eneralized **L**inear **M**odel
GS	**G**rowing **S**eason
IOD	**I**ndian **O**cean **D**ipole
IODZM	**I**ndian **O**cean **D**ipole **Z**onal **M**ode
IPCC	**I**ntergovernamental **P**anel on **C**limate **C**hange
ITCZ	**I**nter **T**ropical **C**onvergence **Z**one
JJA	**J**une **J**uly **A**ugust
MJO	**M**adden **J**ulian **O**sciallation
NCDC	**N**ational **C**limatic **D**ata **C**enter
NCEP	**N**ational **C**enters for **E**nvironmental **P**rediction

NINO	El-**Niño**
NOAA	**N**ational **O**ceanic and **A**tmospheric **A**dministration
OGCM	**O**cean **G**eneral **C**irculation **M**odel
PCA	**P**rincipal **C**omponent **A**nalysis
PCMDI	**P**rogramme for **C**limate **M**odel **D**iagnosis and **I**ntercomparison
PDO	**P**acific **D**ecadal **O**scillation
PP	**P**erfect **P**rognosis
PPE	**P**erturbed **P**hysics **E**nsembles
QBO	**Q**uasi **B**iennal **O**scillation
RCM	**R**egional **C**irculation **M**odel
RHUM	**R**elative **HUM**idity
SA	**S**outhern **A**frica
SAM	**S**outhern **A**nnular **M**ode
SAR	**S**econd **A**ssessment **R**eport
SD	**S**tatistical **D**ownscaling
SHT	**S**outhern **H**emisphere **T**emperature
SLP	**S**ea **L**evel **P**ressure
SST	**S**ea **S**urface **T**emperature
SWIO	**S**outh **W**est **I**ndian **O**cean
TAR	**T**hird **A**ssessment **R**eport
TTT	**T**ropical **T**emperate **T**rough
WCRP	**W**orld **C**limate **R**esearch **P**rogramme
WG	**W**eather **G**enerator

Chapter 1

Background

1.1 Introduction

Since the conclusion of the Intergovernmental Panel on Climate Change (IPCC) Second Assessment Report (SAR), that 'the balance of evidence suggests that there is a discernible human influence on global climate' (Houghton et al., 1995), there has been increased interest in studying the changes in the climate system and its impacts on the environment, economies and societies all around the globe, and in attributing those changes to causes other than natural variability (e.g. anthropogenic causes).

Certain regions and areas have been identified as particularly prone to adverse effects of changes in climatic variables such as rainfall and temperature. According to Solomon et al. (2007), there is high confidence that Africa, and the sub-Saharan region in particular, is probably the most vulnerable continent to climate change (Slingo et al., 2005). The high sensitivity of the region to climate change is aggravated by multiple stresses such as widespread poverty, recurrent droughts and floods. Many other factors, such as heavy disease burden and numerous conflicts, also contribute to reduce the ability of the region to cope with the potential impacts of change. Moreover, according to many studies, global warning may cause intensification of the hydrological cycle, with increases either in the frequency and magnitude of heavy precipitation events, as well as changes in frequencies of extended dry and wet spells (Fauchereau et al., 2003; Huntington, 2006; Williams et al., 2007) with associated scarcity in water availability and drought in parts

of the globe. The hydrological cycle is strictly linked to all aspects of the climate, and changes in it affect water availability through several mechanisms including changing precipitation patterns, intensity and extremes; melting of snow and ice; increasing atmospheric water vapour; changes in evapotranspiration; and changes in soil moisture and runoff (Bates et al., 2008). Furthermore, in the last IPCC report (AR4; Solomon et al., 2007) it is anticipated that more extreme droughts and frequent flooding will strongly affect rainfed agriculture (upon which Africa is heavily dependent), along with health, water resources, ecosystems resilience and infrastructures (IPCC, 2007). The natural hazards connected with changing precipitation patterns are also projected to affect the quality of water and exacerbate many forms of water pollution which, in turn, will affect the availability, stability, access and utilisation of food (Bates et al., 2008).
Thus, the ability to predict the variability and changes in the water cycle and to assess significant changes in rainfall characteristics, at daily and monthly to seasonal timescales, as well as the investigation of its impacts on relevant sectors, is a key requirement if societies are to develop effective adaptation and mitigation responses (Allan and Soden, 2008).

1.1.1 Water resources in southern Africa

Although, as noted in the previous section, there is confidence that water resources in sub-Saharan Africa will become more precarious due to climate change, evidence for historical drying in southern Africa throughout the 20^{th} century is considered to be inconclusive (Mason and Tyson, 2000). Over the past century, a downward rainfall trend, even if not statistically significant, has been reported over the subcontinent (Solomon et al., 2007, Chap. 3, Fig. 3.13 and 3.14). Moreover, the 20^{th} century has been characterised by a high degree of rainfall decadal variability, especially since the 1960s (Giannini et al., 2008), as well as by a number of severe droughts (during the 1980s and 90s).
Future rainfall projections for southern Africa are also not coherent. In a study by Lobell et al. (2008), in order to prioritise climate change food-insecure regions, temperature and precipitation changes were considered as climate variables and southern Africa was

identified as a 'climate risk hot spot' and, with southern Asia, one of the two most vulnerable regions to climate change impacts. However, according to the last IPCC report (Solomon et al., 2007, Figures 10.12 and 11.2) future projections of precipitation in sub-Saharan Africa are characterised by a high level of uncertainty, with lack of agreement on the direction of potential changes (Hulme et al., 2001; Bates et al., 2008; Giannini et al., 2008).

1.2 Southern African climate and rainfall variability

Southern Africa, considered in this study as the region bounded by the Equator and the southernmost point of Africa (Cape Agulhas) at almost 35°S, is characterised by a high degree of climate and rainfall variability on a range of different time scales: inter-annual, inter-decadal and multi-decadal (Mason and Jury, 1997; Todd and Washington, 1999; Dilley, 2000; Reason et al., 2004, 2006). See Mason and Jury (1997) for an extensive review. Scientists currently face major challenges in understanding and predicting this variability (Hulme et al., 2001).

The subcontinent is affected by several distinct large-scale circulation systems: equatorial (from 0° to around 15°S), tropical (coastal desert-type in the Namib, sub-tropical-type over the south-eastern coasts, sub-arid tropical-type on the central plateau) and temperate (Mediterranean-type in Western Cape Province) (Reason and Rouault, 2002; Fauchereau et al., 2003). The large-scale circulation structure is also affected by the orography, with the southern and eastern African plateaus characterised by a 1000m mean and reaching more than 3000m. Steep topography near the western and south-eastern coasts leads to strong climatic gradients and a variety of topographically forced weather systems. Vegetation feedbacks also play an important role on the climate, acting as a focal point for land surface-atmosphere interactions (Foley et al., 1998).

The region is influenced by tropical, subtropical and midlatitude pressure systems. The quasi-stationary anticyclones, corresponding to the descending branch of the Hadley cells, constitute the semi-permanent subtropical high pressure belt which varies in position throughout the year following a semi-annual oscillation (SAO), arising from differential warming between the Southern Ocean and the Antarctic Continent. The neighbouring oceans influence both temperature and moisture supply over the region, modulating

climate and atmospheric features (Mason, 1995; Tyson and Preston-White, 2000; Vigaud et al., 2008; Giannini et al., 2008). Along the west coast the southern African climate is affected by the cold Benguela current, where superficial water is transported away from the coast driven by the prevailing south-easterly trade winds. The sea surface is depressed near the coast and an upwelling system with waters below 15°C occurs in compensation (Tyson and Preston-White, 2000). In contrast along the east coast the warm water of the western boundary of the Agulhas current, with temperature between 22°and 28°C in summer, flows adjacent to the coastline.

In the equatorial region the equatorial current system consists of the North Equatorial Current (NEC), flowing westward around 20°N, the South Equatorial Current (SEC) flowing westward around 0° to 5°S and, and between them, the Equatorial Counter Current (ECC) flowing eastward around 10°N (Gasse et al., 2008). The Counter Currents result from balancing the westward flow of water in each ocean by the North and South Equatorial currents. In the Indian Ocean the SEC is driven by the South Indian Anticyclone in the atmosphere. The portion of current overtaking Madagascar from the north bifurcates turning northwards and southwards after rounding the northern tip of the island. The water flowing southwards through the Mozambique Channel contributes to the Agulhas Current to the south. The portion of SEC turning south after reaching the coast of Madagascar, mainly turns backwards towards the South Indian Ocean (Tyson and Preston-White, 2000). South of 40°C the Antarctic circumpolar current (ACC) serves as a principal pathway of exchange between the South Atlantic and the South West Indian Ocean, transporting huge quantities of cold water eastwards around the globe. From the east the Agulhas current bends backwards becoming the Agulhas Return Current, while on the west the South Atlantic Anticyclone drives the Benguela Current offshore. Consequently the southern Cape is characterised by intensive turbulence and mixing of water, and is one of the most variable regions of the world oceans (Tyson and Preston-White, 2000). The climatic gradients due to the described features of the subcontinent lead to zonal and meridional division of rainfall.

As shown in Figure 1.1, produced using the CRUTS3 data set described in Section 2.2, highest precipitation is found in the equatorial region, with peaks over 200mm/month, and, during the austral summer, along a dipole structure defined as tropical-temperate trough (TTT) where a gradient of precipitation from northwest to southeast is reflected

in the monthly mean precipitation (Harrison, 1984; Todd and Washington, 1999; Washington and Todd, 1999).

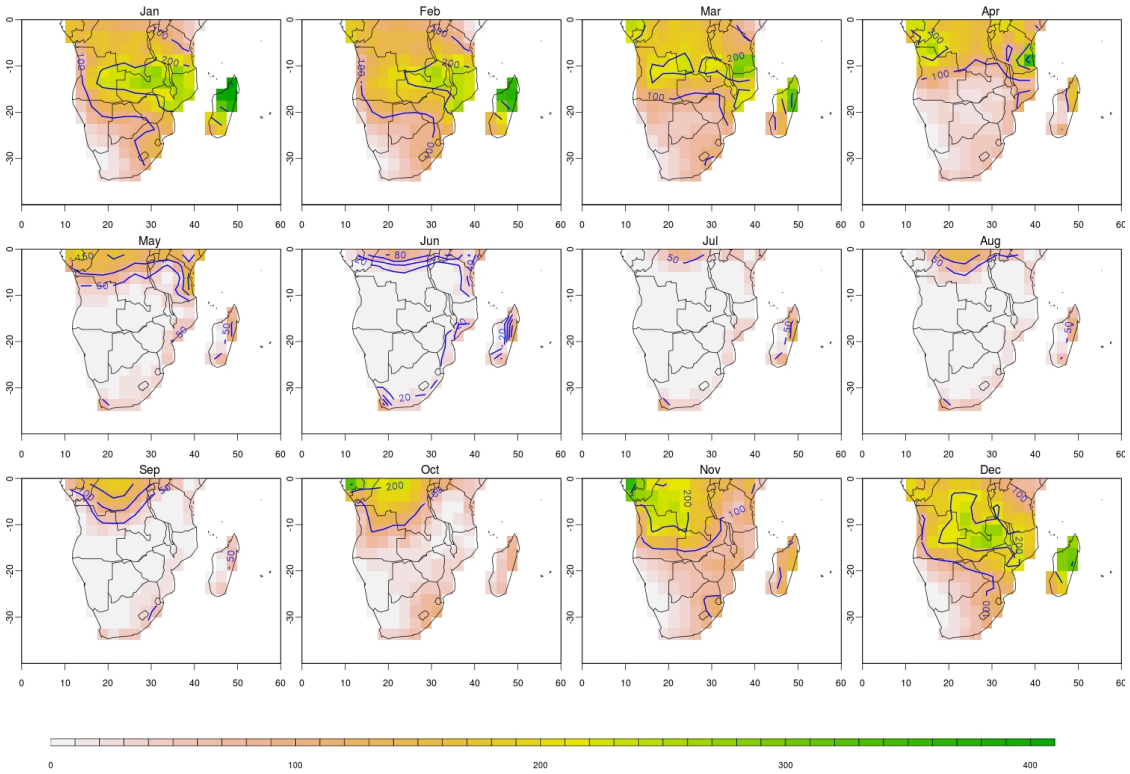


FIGURE 1.1: Monthly precipitation mean (mm) for the period 1951-2000.

The TTTs form when a tropical low (usually over Angola) is coupled to a wave perturbation in the temperate westerlies (Mason and Jury, 1997) linking the tropical convection to the midlatitude weather systems. This major mechanism of poleward transfer of energy and momentum in southern Africa (Todd and Washington, 1999) becomes stationary reflecting the mean position of the TTTs.

The most important source of variability and principal rainmaking mechanism in southern Africa is the migration of the Intertropical Convergence Zone (ITCZ), an area of low pressure formed where the Northeast Trade Winds meet the Southeast Trade Winds near the Equator. As these winds converge, moist air from the tropical oceans is forced upward into the equatorial trough zone, a quasi-continuous area of low pressure between the subtropical high pressure areas in both the northern and southern hemisphere. This causes water vapour to condense, resulting in a band of heavy precipitation around the globe. Changes in air pressure over land cause a seasonal shift in the location of the

ITCZ. Over the African continent the ITCZ shifts significantly from January to July causing a pronounced seasonal cycle with wet (from around November to April) and dry seasons across southern Africa. The south-westernmost Cape is an exception, with most rainfall occurring in the austral winter (maximum around June).

According to Van Heerden and Taljaard (1998) other major features of surface circulation during the austral summer (mainly DJF) are: the northeast monsoon system, which crosses the Equator and spreads into east Africa and southwards to the ITCZ; the southerly trade winds between the Atlantic anticyclone and the subcontinent becoming the southwest monsoon entering the land between 12°S and 5°N; the east-west feature along the southern branch of the ITCZ indicated as Congo Air Boundary (CAB) or Inter-Ocean Convergence Zone (IOCZ) representing the boundary between the dry continental southeast trade winds and the moist southwest monsoon air; the westerlies over the Congo basin which extend eastward to the ITCZ in the north and southward to the IOCZ in the south; the easterly trades over Zimbabwe recurving anticyclonically southward over Botswana and Namibia; the westerly current over the extreme southwestern plateau of South Africa, which splits to become either southerly or easterly over the plateau.

The seasonal variation in pressure and wind fields is much more dramatic in the Indian Ocean than in the Atlantic. The pressure on the South African east coast rises (by about 10 hPA) due to the northward shifts of the ITCZ and the northeast monsoon is replaced by a southerly monsoon consisting of a southeasterly trade wind changing direction close to the Equator and becoming the southwest monsoon. Over land the remains of the IOCZ lies above the Congo basin, where the Atlantic maritime air penetrates and, due to the pressure rise, the circulation over much of the southern African plateau reverts into a constant southeasterly to easterly trade-wind flow.

1.2.1 Controls of southern African rainfall variability

After having introduced the southern African climate, the current section provides an insight on the literature about the relationships and drivers of the subcontinental rainfall variability. Such investigation is subsequently used during the selection of the potential predictors to be employed in the modelling exercise presented in Chapters 3, 4 and 5.

This rainfall variability has been explained in terms of changes in large-scale weather systems, their frequency, duration and intensity (Mason and Jury, 1997) and variations in pressure and atmosphere circulation patterns (Tyson, 1981).

Many studies have emphasized the influence of sea surface temperatures (SSTs) on southern African rainfall variability. In particular, El Niño Southern Oscillation (ENSO) is considered the dominant interannual mode of variability in the tropics with a clear marked influence upon interannual rainfall variability over the subcontinent (Nicholson and Kim, 1997; Mason, 2001; Reason and Jagadheesha, 2005a; Giannini et al., 2008). ENSO is a global coupled ocean-atmosphere phenomenon with an ocean signature of SST fluctuations (El Niño and La Niña in the tropical eastern Pacific Ocean) and an atmospheric signature, the Southern Oscillation (SO), which reflects the monthly or seasonal fluctuations in the air pressure difference between Tahiti and Darwin. During El Niño events and Southern Oscillation Index low phase, convection is suppressed over Indonesia with an eastward shift of the ascending limb of the Walker circulation over the western Indian Ocean; this leads to a reduced meridional flux of energy over southern Africa, while convection is increased over Madagascar through the eastward shift of the cloud-band convergence zones. Due to its effect on the location of major cloud bands over southern Africa, large scale warming of the equatorial eastern and central Pacific is frequently associated with drought over much of the subcontinent (Mason and Jury, 1997; Mason, 2001; Richard et al., 2001; Jury et al., 2004; Tyson and Preston-White, 2000; Reason and Jagadheesha, 2005a; Giannini et al., 2008; Schubert et al., 2009; Findell and Delworth, 2010). In particular, El Niño impacts are thought to be most pronounced in the south-east of the continent (Matarira, 1990; Rocha and Simmonds, 1997b) during the austral summer (Tyson and Preston-White, 2000). From January to March significant correlation between the El Niño signal and rainfall has been shown by Nicholson and Kim (1997) through a composite analysis of rainfall during El Niño events. The same link is confirmed by several other studies (among which Lyon and Mason 2007, 2009, via composite analysis and model experiments). However, Rocha and Simmonds (1997a) found, analysing the correlation between southern African summer rainfall and different lead times in the SOI index, the strength of the association to vary. Good examples of the significant variation in the impacts of El Niño over southern Africa occurred in the last two decades when during relatively weak 1992/3 and 2002/3

El Niño events southern Africa experienced widespread summer droughts. On the other hand, dry conditions were less intense during one of the strongest events on record in terms of the Southern Oscillation Index and SST anomalies in the tropical Pacific, in 1997/8 (Reason and Jagadheesha, 2005a). Therefore, ENSO appears to be not the only component affecting rainfall over southern Africa and other factors may contribute in driving its variability.

In addition, positive correlations have been found between southern African summer rainfall and SST anomalies around the subcontinent, in both the Benguela Current system along the west coast (correlation range 0.31/0.35) and the Agulhas Current in the Indian Ocean (correlation range 0.35/0.40; Walker 1990). With warm waters, the Indian Ocean is considered as the major source of moisture over the subcontinent. Hoerling et al. (2006) linked warming and enhanced convection over the entire Indian Ocean with subsidence and drought over southern Africa during the austral summer months. With the use of GCM simulations they showed a rainfall reduction of as much as 100 mm across the subcontinent to a specified 1°C Indian Ocean sea surface warming, during the February-April season.

Washington and Preston (2006) identified a dipole pattern in the South West Indian Ocean (warmer conditions in the south-west and cooler in the north-west) through the use of a GCM idealized SSTs experiment. In the study, an increased SST gradient between a warm anomaly centred at 32°S, 55°E and cold anomaly centred at 12°S, 65°E was associated with extremes in observed southern African austral summer rainfall (positive rainfall anomalies around 1-4 mm/day). Reason (2001) linked the east-west Indian Ocean Dipole Zonal Mode (IODZM; Saji et al. 1999) to increased moisture advection over southern Africa when SSTs increase from east to west. Moreover, Manatsa et al. (2008) analysed the correlation between a standardised Zimbabwean precipitation index and basin wide temperature anomalies over the tropical Indian and Pacific Ocean for the period September-November, represented via the IODZM index. The index appeared to have a stronger impact on Zimbabwe's rainfall than El Niño when the two are in competition, and to be associated with rainfall variability.

Recent investigations have also considered the Atlantic Ocean as a source of southern African rainfall variability, both for south-west Cape winter rainfall and for summer rainfall anomalies particularly over Angola and Namibia (Reason and Jagadheesha, 2005b;

Reason et al., 2006). Williams et al. (2008), investigating the association between southern African rainfall and atmospheric and oceanic circulations, identified a combination of anomalously cold SSTs in the central South Atlantic and warm SSTs off the south-western coast associated with an increase in rainfall extremes over the subcontinent.

Other large-scale features of the climate system have been associated with variations in the southern African rainfall regime. The Brandon-Marion Index (BMI), indicative of changes in the pressure field over the South West Indian Ocean, was found to be positively correlated with south-eastern African summer rainfall (Rocha and Simmonds, 1997a). In particular they found the highest correlation in an area north-east of South Africa during the summer months (Dec-Jan).

The Quasi-Biennial Oscillation (QBO), a quasi-periodic oscillation of the equatorial zonal wind between easterlies and westerlies in the tropical stratosphere with a mean period of 28 to 29 months, is thought to modulate the effect of ENSO over southern African rainfall (Jury et al., 1994). It is suggested that lower stratospheric easterly zonal winds may enhance Walker cell overturning with a descending limb over southern Africa, whereas during westerly phase years, the reversal of the Walker cell could generate a rising limb over southern Africa and enhance convection and rainfall over the subcontinent (Mason and Jury, 1997).

The Southern Annular Mode (SAM; Rogers and van Loon, 1982), also known as the Antarctic oscillation (AAO), is a zonally-symmetric model of tropospheric circulation variability in the southern hemisphere (south of 20°S) with opposing geopotential height perturbations of opposite signs over the Antarctic and a zonal band centred near 45°S. Reason and Rouault (2005) found a relationship between the positive phase of the SAM and drier winter conditions over southwestern South Africa. Correlating the time series of the SAM index and a South African rainfall index they found values as high as 0.4. On the other hand, the positive phase of the SAM has been associated with anomalously wet conditions over much of the rest of South Africa, with an increase in precipitation around 0.5mm/day (Gillett et al., 2006).

The Madden-Julian oscillation (MJO), a mode of intraseasonal variability that affects the location and strength of tropical precipitation, is characterised by an eastward propagation of tropical deep convection clusters from the Indian Ocean to the western Pacific Ocean. The MJO was found to significantly influence rainfall in the east and south of

southern Africa (Pohl et al., 2007).

The North Atlantic Oscillation (NAO), a phenomenon in the North Atlantic Ocean characterised by fluctuations in the difference of sea-level pressure between the Icelandic Low and the Azores High, was linked by McHugh and Rogers (2001) to precipitation variability along and north of the southeastern Africa convergence zone. In the study they described anomalously high (low) convective rainfall occurring over southeast Africa when the NAO is weak (strong).

Besides variability at the intraseasonal and interannual time scales, Reason and Rouault (2002) linked the Pacific Decadal Oscillation (PDO), a pattern of oscillation in the Pacific Ocean that shifts phase on an interdecadal time scale, to South African rainfall. They noticed that the warm (cool) phase of the interdecadal variability in the Pacific and Indian Ocean was associated with decreased (increased) rainfall over South Africa (correlation coefficients range -0.4/-0.53).

In the Atlantic Ocean the Atlantic Multidecadal Oscillation (AMO) is a near-global scale mode of observed multidecadal climate variability with alternating warm and cool phases over large parts of the Northern Hemisphere. Ward (1998) linked the AMO to the multidecadal dry condition over sub-Saharan Africa.

Finally, the displacement and change in strength of the subtropical jet stream, a belt of strong upper-level winds lying above regions of subtropical high pressure, was related to rainfall variability over the subcontinent by Richard et al. (2001). According to the study, a strengthening and a northward displacement of the jet stream could lead to a displacement of the cyclonic systems tracks further north moistening the southern latitude of the continent, a decreasing of the easterly wind component and a reduction in tropical temperate troughs over central southern Africa.

1.3 General circulation models and associated uncertainty

One of the aims of this research is to examine future projections of rainfall over southern Africa at different temporal and spatial scales. Nowadays, most such projections are based on the outputs of the three-dimensional climate models, conventionally known as general circulation models (GCMs), which are extensively reviewed in the second (SAR; Houghton et al., 1995), third (TAR; Houghton et al., 2001) and the fourth assessment

reports (AR4; Christensen et al., 2007). The IPCC AR4 states that *'there is considerable confidence that Atmosphere-Ocean General Circulation Models (AOGCMs) provide credible quantitative estimates of future climate change, particularly at continental and larger scales'* (Christensen et al., 2007, Chap. 8).

GCMs are mathematical representations of the climate system expressed as computer codes, developed and used to project plausible projections for the coming century and beyond. Alternative scenarios, particularly with reference to the production of greenhouse gases and aerosol precursor emissions, are introduced to explore future developments in the global environment. These scenarios are based on relationships between key driving forces such as demographic change, social and economic development, and the rate and direction of technological change. The IPCC developed 40 different scenarios classified into 4 scenario families or storylines. The Special Report on Emissions Scenarios (IPCC, 2000) defined those storylines describing the relationships between the forces driving greenhouse gas and aerosol emissions and their evolution during the 21^{th} century for large world regions and globally (A1, A2, B1 and B2). Each scenario makes different assumptions for future greenhouse gas emissions, land-use and other driving forces as products of hypothesis about future technological, social and economic development that diverge in increasingly irreversible ways.

Those future scenarios allow the study of the effects of changes in climate variables on other systems, sector and region of the globe (e.g. hydrology and water system, agricultural sector, human health). Although there are methodological problems linked to the application of scenarios, such as the downscaling of social, economic and technological development projections, they still provide a coherent global quantification of future development and future emissions available for impact studies (Parry et al., 2007).

1.3.1 Climate Model structure

GCMs are mathematical models of the general circulation of the planetary atmosphere and/or ocean and they can generally be split in three categories:

a) Atmosphere general circulation models (AGCMs) consist of a three-dimensional representation of the atmosphere coupled to the land surface and cryosphere. The

AGCM has to be provided with data for sea-surface temperatures and sea-ice coverage. Hence, an AGCM by itself cannot be used for climate prediction, because it cannot indicate how conditions over the ocean will change. AGCMs are useful for studying atmospheric processes, the variability of climate and its response to changes in sea-surface temperature;

b) Ocean general circulation models (OGCMs), the ocean counterpart of an AGCM, are three-dimensional representation of the ocean and sea ice. OGCMs are useful by themselves for studying ocean circulation, interior processes and variability, but they depend on being supplied with data about surface air temperature, winds and other atmospheric properties;

c) Coupled atmosphere-ocean general circulation models (AOGCMs) are the most complex models in use, consisting of an AGCM coupled to an OGCM. Some recent models include the biosphere, carbon cycle and atmospheric chemistry as well. AOGCMs can be used for the prediction and rate of change of the future climate. They are also used to study the variability and physical processes of the coupled climate system.

All GCMs attempt to represent the atmosphere/ocean dynamics by solving the relevant mathematical equations of fluid flow. In practice, approximate solutions must be found on a three-dimensional grid, the resolution of which is limited by the available computing power (current global climate models typically have a grid resolution of a couple of hundred kilometres at the earth's surface). Moreover, parametrizations are used to include the effects of various sub-grid scale processes that have dimensions smaller than the model resolution or processes that are too complex to be represented. In such cases parameters replace equations explicitly resolving processes. Examples include cloud convection, cloud microphysics and the radiative transfer process.

Although AOGCMs provide the most comprehensive information currently available about the global climate system, and form the basis for most projections of future climate, the interpretation of such projections must acknowledge that the climate is a complex chaotic system. As such, the best that any model can hope to do is to produce a sequence which is, in terms of its properties, indistinguishable from the true trajectory (at least within the period of interest) of the variable being considered into account (Leith, 2005a).

It is important to highlight that several sources of uncertainty, the so called 'uncertainty cascade', are associated with the construction and application of climate models and use of their outputs. These sources can be broadly classified into forcing scenario, initial conditions, parametrization and model structure uncertainty (see next section).

1.3.2 Sources of uncertainty

One of the difficulties in making long-term climate projections is that, due to the long lead times involved, there is no possibility of a calibration for the forecast regime of interest (Stainforth et al., 2007). This contrasts with short-term weather forecasting, where accumulated experience can be used to improve forecast performance.
Although there have been huge advances in the complexity and realism of GCMs since the first models were developed in the 1960s, several sources of uncertainty remain in the projections obtained from these models (Giorgi and Francisco, 2000; Stainforth et al., 2007; Collins, 2007). The 'cascade of uncertainty' can be defined as follow:

a) Forcing scenario uncertainty, that involves factors outside the realm of the climate but that affect it. This uncertainty is due to the global system unpredictability and it includes factors such as greenhouse gases, aerosols, solar and volcanic forcing. Since a formal quantification of future anthropogenic forcing agents has been said to be impossible (Collins, 2007), the idea that we will ever be able to produce probabilistic predictions with all uncertainties quantified has been dismissed by many;

b) Inter-model uncertainty, also referred as 'model uncertainty' by Stainforth et al. (2007), which leads to disagreement between GCMs regarding the magnitude and sometimes the direction of future change. Indeed, different GCMs exhibit varying levels of performance over different regions and for different climatic variables (Christensen et al., 2007). This is mainly due to different representations of physical processes, grid resolutions and numerical schemes used to solve the dynamical equations and the choice of processes parametrization and feedbacks of climate (e.g.

cloud formation, cloud and cryosphere feedbacks and land surface effects);

c) Initial condition uncertainty due to the imprecise knowledge of the current state of the system, such as temperature, pressure and humidity among others. Since the climate system is chaotic, small differences in model fields grow over time potentially leading to big differences in the future projections. This source of uncertainty also arises from the quality and reliability of historical data, observations, and measurements from which the initial conditions are set;

d) Model inadequacy, arising from the fact that any model is at best an approximation of reality.

Moreover, in case of impact studies applications it is also important to include the uncertainties in the impact models themselves.

1.4 Ensembles of simulations

Given the inter-model uncertainty introduced in the previous section and the differences between rainfall projections from different GCMs in areas such as southern Africa (Solomon et al., 2007, Chap.11, Fig.11.2), it is extremely difficult to identify the most reliable model simulation and impossible to recognize a 'true' climate model. As a consequence, it is increasingly being recognised that a useful interpretation of such climate models' outcomes must confront the inherent uncertainty. Climate assessments should be based on the combined information provided by ensembles of different GCM simulations, rather than just a single model or multi-model mean (Räisänen and Palmer, 2001).

A single-model ensemble involves the use of a number of realisations of a single deterministic model. Distinct predictions are obtained for each realisation by either perturbing the initial conditions that are all plausible given the past and current set of observations and varied by an error whose magnitude reflects the degree of uncertainty, or by selecting

different sets of model parameters (Viney et al., 2005). These ensembles are defined as 'initial condition' and 'perturbed physics' ensembles (PPE), whereas in a multi-model ensemble, several different deterministic models are used simultaneously. A combination of both may also be used. A recently introduced possibility is the so-called stochastic parametrization or stochastic physics, where random perturbations to sub-grid-scale unresolved processes are included (Collins, 2007).

As pointed out by Tebaldi and Knutti (2007), a single PPE approach is limited in its ability to capture the full range of uncertainties in the models' representation of the climate system, as there are many ways to design a parametrization and it is expected that the choice of a single model run within a PPE would not produce significantly different outcomes from an alternative choice within the same ensemble, when summary statistics of climate-driven results are computed (typically averaged over decades). Therefore the quantification of all aspects of model uncertainty requires multi-model ensembles, found generally to give wider range of sampled uncertainty than single-model forecasts. Several studies have been made to combine models' output and quantify inter-model uncertainties: based on expert opinions (Morgan et al., 2001), by comparing different outputs from several GCMs in different regions (Giorgi and Francisco, 2000), by taking a simple multimodel ensemble average (Giannini et al., 2008), by weighting the models (Tebaldi et al., 2005) or by a probabilistic interpretation of a multimodel ensemble for climate projections through different approaches (Dessai et al., 2005).

1.5 Scaling GCM outputs

Substantial progress has been made in modelling the climate in the latest years: increasing the resolution, improving the parametrization and adding further processes in most of the models. Nonetheless, up-to-date GCMs do not provide reliable information below around 200km of resolution (Meehl et al., 2007) and important deficiencies still remain in the simulation of clouds and tropical precipitation. One reason for such deficiencies is ascribable to the fact that small-scale process are not directly resolved by numerical equations, rather approximated via parametrizations. The deficiencies in simulating clouds are especially relevant for the tropical areas (Solomon et al., 2007). The problem

here is due both to the sub-grid scale of the processes, when a partial cloud coverage can occur due to subgridscale fluctuations of humidity and temperature, and to the limitations in the scientific understanding of the cloud formation (Christensen et al., 2007). As a result it is still not advisable to make direct use of precipitation as projected by the GCMs in impact study analysis, where hydrological processes occurring at a finer temporal and spatial scale are of interest.

In order to overcome such limitations, downscaling techniques attempt to resolve the scale discrepancy between such climate change scenarios and the resolution required for impact assessments (Wilby and Wigley, 1997; Maraun et al., 2010). The methodologies used can be summarized in two broad categories: dynamical downscaling, involving the explicit solving of the physical dynamics of the system, and statistical downscaling, identifying relationships between circulation and the local climate (Hewitson and Crane, 1996). A number of previously published papers have reviewed the different techniques and advantages and disadvantages of both approaches (Hewitson and Crane, 1996; Wilby and Wigley, 2000; Wilby et al., 2004; Fowler et al., 2007; Maraun et al., 2010).

Dynamical downscaling includes the use of a regional climate model (RCM) nested in a GCM from which to derive larger-scale information. Such models are computationally expensive and the development and application of high-quality RCMs over the study region is still limited. For such reasons this technique is not reviewed further in the present study.

1.5.1 Statistical downscaling

As an alternative approach, statistical downscaling is based on the view that the local climate is conditioned by the large scale state of the climate and the regional physiographic features. Thus, a statistical model relates the large-scale climate variables to regional and local variables in order to estimate the corresponding local climate characteristics (Wilby et al., 2004). Statistical downscaling consists of identifying empirical links at a specific location between large-scale patterns of climate elements (predictors) and local climate (the predictand), and applying them to output from global or regional models. Successful statistical downscaling is therefore dependent on long reliable series of predictors and predictands (Draggan, 2010).

Although there is a general consensus that the most appropriate choice of predictor variables should include measures of temperature, atmospheric circulation and moisture, there is less agreement on how to specify them (Fowler et al., 2007). However, Wilby and Wigley (2000) extensively reviewed the procedures for choosing the most suitable predictor variables.

Some assumptions need to be taken into consideration when using statistical downscaling techniques. Wilby et al. (2004) and Leith (2005a) summarized them in three categories:

a) as already mentioned, the climate variable of interest should vary in response to large scale atmospheric structure according to the local, large scale and time interval considered as well as atmospheric variables taken into account;

b) the large scale predictors used in the statistical model must respond to greenhouse gases forcing in the climate model and be correctly represented by the climate model. However, this assumption can never be verified, since we cannot necessarily assume that a good present climate performance reflects a good performance under a future climate scenario;

c) the statistical relationships between the local variable and the large scale structure developed under the present day climate must remain valid under future climate conditions (Wilby et al., 2004). Of course, the same assumption also applies to the parametrizations used in RCMs, although the physically based components of the model ensure that the problem is less severe here.

Several classifications have been proposed to divide the statistical downscaling methods. In general they are categorised into regression methods, weather typing approaches and stochastic weather generators (WGs; Wilby and Wigley 1997; Fowler et al. 2007). Rummukainen (1997) classified them into perfect prognosis (PP), if the predictand is related to observations or representations of them at (nearly) concurrent times, and model outputs statistics, which establish statistical linkages directly between the climate model outputs and the real world observations. For a more comprehensive review about all such methods, refer to Wilby and Wigley (1997); Fowler et al. (2007); Maraun et al. (2010).

The technique used in the present research is an hybrid between a WG, a statistical model

that generates local scale time series resembling the statistical properties of observed weather, and a PP statistical downscaling method, conditioning its parameters on large scale predictors (Maraun et al., 2010). Such statistical framework was initially employed to relate large scale climate features to regional weather in order to investigate the potential controls on southern Africa rainfall variability (Chapter 3), and subsequently applied to downscale GCM outputs at subcontinental (Chapter 4) and finer spatial scales (Chapter 5).

1.6 Research objectives and thesis outline

From the foregoing discussion it is clear that southern Africa is affected by a high level of rainfall variability at different time scales. Such variability exacerbates the multiple stresses already present, making the region even more sensitive to possible future impacts of climate change. State-of-the-art GCMs, although providing valuable climate information under future scenarios, are still subject to limitations and sources of uncertainty. Moreover, because of their relatively coarse spatial scale, GCMs are mostly inadequate to assess the local-scale impacts of changes in climate. Therefore, in order to prioritise adaptation and mitigation measures, further research is required to characterise rainfall variability and translate such information to scales that are relevant for impact studies.

Therefore, the purpose and principal aims of this study are:

- To characterise the southern African precipitation climatology and variability and to examine performance and limitations of GCMs in reproducing it (Chapter 2);
- To investigate the relationships between southern African monthly rainfall and large-scale climate factors, building upon current knowledge of the structures controlling subcontinental rainfall variability (Chapter 3);
- To develop a tool for simulating probabilistic rainfall data sets, conditional on the identified climate variables, for the investigation of future scenarios of monthly regional rainfall, without the direct use of GCM precipitation (Chapter 4);

- Finally, to assess the suitability of the developed model for impacts applications in agriculture. Specifically, the aim is to use the previously identified large-scale atmospheric and oceanic variables to downscale daily precipitation for a narrower study area from which indices relevant for the characterisation of the maize growing season are computed. Outputs from the study may be used to inform local farmers in addressing management practices and mitigation actions (Chapter 5).

In Chapter 2 the data used and preliminary analysis on GCM limitations in representing regional precipitation are presented. A technique is developed in Chapter 3 to characterise the subcontinental monthly rainfall variability without the direct use of climate models' precipitation. The used framework simultaneously links several large-scale climate features to regional climate. The stochastic models are fitted to observed data and used to generate synthetic weather time series to investigate the regional monthly rainfall variability.

In Chapter 4 the developed framework is used to downscale the future projections of regional precipitation under a chosen emission scenario. Finally, a similar methodology is used to study daily (rather than monthly) precipitation characteristics for a specific local-scale case study in South Africa (see Chapter 5). In particular, rather than only studying the seasonal rainfall totals, a special focus is given to changes in rainfall characteristics within the maize growing season to illustrate a direct application of the developed framework in an investigation of climate change implications for maize cropping.

The final chapter provides a synthesis of the conclusions from the present work and a brief discussion into possible directions for future research.

Chapter 2

Preliminary analysis

2.1 Introduction

As stated in Section 1.3, GCMs are one of the primary tools in projecting future climate scenarios. However, it is widely known that they have some limitations and, since no model reproduces the 'true' climate, different models provide different results (Weaver and Zwiers, 2000).

This chapter presents a thorough preliminary exploration of both rainfall observational data and GCM projections in qualitative terms.

The work presented here does not attempt to quantify exactly the ability of GCMs in representing the observed climate. Rather it provides a foundation for the more rigorous analysis that follows (Chandler and Scott, 2011, Ch. 2). In particular, the aims of these analyses were threefold: first, to characterise the precipitation climatology of the study region as regards aspects other than just mean rainfall (which has been discussed extensively in the literature review; see Section 1.2); second, to gain a preliminary understanding of the similarities and differences between GCM-simulated and observed precipitation fields for the future modelling; and finally to establish a firm basis for the subsequent statistical modelling exercise reported in Chapters 3, 4 and 5.

2.2 Data

Southern Africa is here defined as the area lying between -40° – 0°N and 0° – 60°E. To investigate the 20th century climate of the region, the observational data set used is the Climate Research Unit TS 3.0 (CRU 2008; from now on referred as CRUTS3) precipitation dataset, with a data coverage spanning 1901–2006. The dataset is an observed gridded land surface precipitation product at monthly temporal resolution and 2.5° × around 2.5° spatial resolution covering the global land surface. It has been created by interpolation from a non regular network of stations from different sources. See Mitchell and Jones (2005) for description of the station sources used to construct the database and procedure in deriving the gridded product. Refer to Section 3.2 for caveats on the use of gridded datasets and motivation for their usage in the present study.

The study also uses data from 23 GCMs, developed as part of the World Climate Research Programme's (WCRP's) Coupled Model Intercomparison Project phase 3 (CMIP3) multi-model dataset (see Table 2.1), and as used in the IPCC Fourth Assessment Report (AR4; Solomon et al. 2007). For each GCM, data from all of the ensemble members (see numbers of runs in Table 2.1) were pooled prior to calculating the considered statistics. When needed, the outputs from different GCMs where regridded to a common 2.5° × 2.5° (corresponding to around 277 km in Latitude × 241 km in Longitude, at 30° S) spatial resolution consistent with the CRUTS3 outputs.

Initial investigations focused on the 100-year period from 1901 to 2000; however, due to restricted availability of data on potential drivers of precipitation the investigation was subsequently concentrated on the second half of the twentieth century (1951–2000). This 50-year period will be the focus of the results reported in the next sections.

The analyses were carried out using the R software environment (R Development Core Team, 2006).

Centre	*Country*	*Model Name*	*Runs*
Beijing Climate Center	China	BCC-CM1	4
Bjerknes Centre for Climate Research	Norway	BCCR-BCM2.0	1
National Center for Atmospheric Research	U.S.A.	CCSM3	6
Canadian Centre for Climate Modelling & Analysis	Canada	CGCM3.1(T47)	5
Canadian Centre for Climate Modelling & Analysis	Canada	CGCM3.1(T63)	1
Météo-France / Centre National de Recherches Météorologiques	France	CNRM-CM3	1
Commonwealth Scientific and Industrial Research Organisation	Australia	CSIRO-Mk3.0	3
Commonwealth Scientific and Industrial Research Organisation	Australia	CSIRO-Mk3.5	3
Max Planck Institute for Meteorology	Germany	ECHAM5/MPI-OM	4
Meteorological Institute of the University of Bonn & Meteorological Research Institute of KMA	Germany / Korea	ECHO-G	5
US Dept. of Commerce / NOAA / Geophysical Fluid Dynamics Laboratory	U.S.A.	GFDL-CM2.0	3
US Dept. of Commerce / NOAA / Geophysical Fluid Dynamics Laboratory	U.S.A.	GFDL-CM2.1	2
NASA / Goddard Institute for Space Studies	U.S.A.	GISS-AOM	2
NASA / Goddard Institute for Space Studies	U.S.A.	GISS-EH	5
NASA / Goddard Institute for Space Studies	U.S.A.	GISS-ER	9
LASG / Institute of Atmospheric Physics	China	FGOALS-g1.0	3
Institute for Numerical Mathematics	Russia	INM-CM3.0	1
Institut Pierre Simon Laplace	France	IPSL-CM4	1
Center for Climate System Research (The University of Tokyo), National Institute for Environmental Studies and Frontier Research Center for Global Change (JAMSTEC)	Japan	MIROC3.2 (hires)	1
Meteorological Research Institute	Japan	MRI-CGCM2.3.2	5
National Center for Atmospheric Research	U.S.A.	PCM	4
Hadley Centre for Climate Prediction and Research / Met Office	UK	UKMO-HadCM3	2
Hadley Centre for Climate Prediction and Research / Met Office	UK	UKMO-HadGEM1	2

TABLE 2.1: GCMs considered in the analysis performed in the current chapter.

2.3 Preliminary analysis

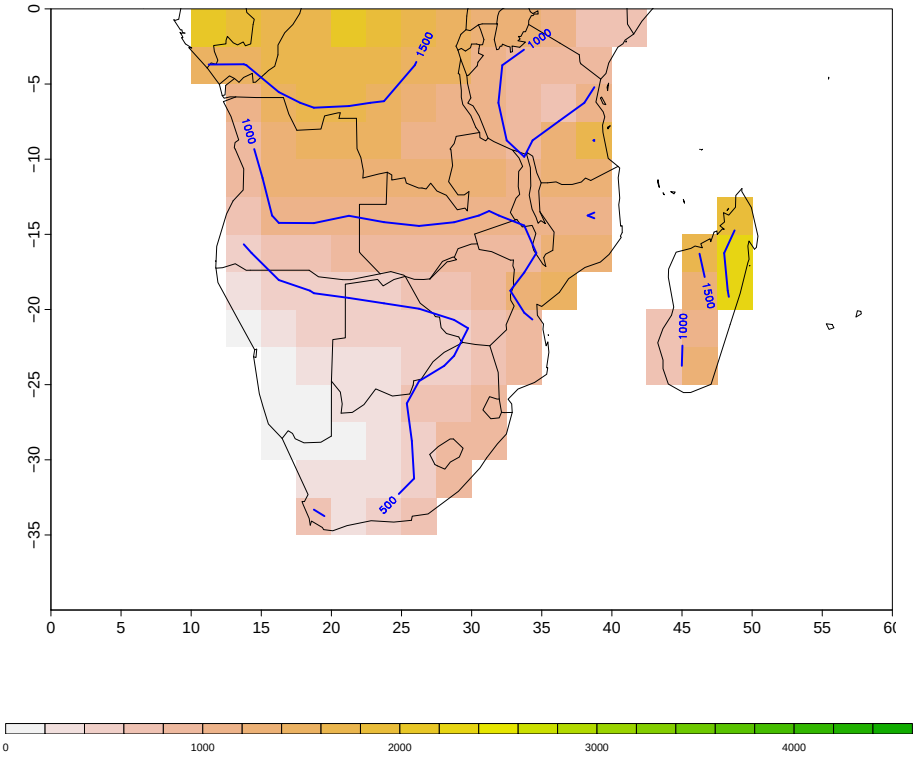
2.3.1 Overall mean precipitation

The first quantity to be investigated was the overall annual mean precipitation at each grid cell for the period 1951–2000. Although it does not provide information about rainfall variability, this metric is useful to characterise the climatology of the region.
The CRUTS3 dataset shows a northwest-southeast gradient over southern Africa (Figure 2.1) that reflects primarily the intense summer rainfall activity in the region (Figure 1.1). Mean precipitation is highest over the Congo Basin where it reaches around 2000 mm/year; over the southeast cape, the precipitation decreases to around 800 mm/year while in the Namib Desert it is well below 200 mm/year.
Figure 2.2 shows the annual precipitation mean modelled by the 23 GCMs investigated. The colour scales are the same as in Figure 2.1.
The patterns of precipitation mean represented by the GCMs are consistent with the observed, with the exception of the Chinese BCC-CM1 (top left corner plot), where the highest values lie, instead, over the equatorial Indian Ocean and the Congo Basin, and the lowest values over the south-west coast and inland (part of Angola, Namibia, South Africa and Botswana). According to Figure 2.2, in general, the models overestimate the average amount of annual rainfall by comparison with the CRUTS3 data, especially over the west equatorial area of the subcontinent (BCCR-BCM2.0, CNRM-CM3, CSIRO-Mk3.0 and 3.5, FGOALS-g1.0, IPSL-CM4, MIROC3.2 (hires), MRI-CGCM2.3.2, NCAR-CCSM3, NCAR-PCM and UKMO-HadCM3). Similarly to the CRUTS3 data some of the models (BCCR-BCM2.0, CGCM3.1(T63 and t47), CNRM-CM3, CSIRO-Mk3.0 and 3.5, GFDL-CM2.0 and 2.1, GISS-AOM, MIROC3.2 (hires), ECHO-G, ECHAM5/MPI-OM, NCAR-CCSM3, NCAR-PCM and UKMO-HadGEM1) show a well defined wetter area over the Kwazulu/Natal representing higher quantities of moisture, presumably driven by the Indian Ocean, on the south-east coast of the continent.
The CRUTS3 dataset also shows a high precipitation mean over the eastern coast of Madagascar, only depicted by a few models. However, the number of interpolated stations used in the gridding procedure for Madagascar was quite low, reducing the reliability of the gridded product over the island.

Figure 2.3 compares the individual ensemble members for each of the GCMs. It shows the overall yearly precipitation mean averaged over the study region for each ensemble member (blue dots). The mean for the observational gridded data is shown as a black solid line. Although the multi-model average of the overall regional precipitation (blue dashed line) is within 10% of the observational value, there is considerable variation among the GCMs, with some of them producing extremely good results (CSIRO-Mk3.0, ECHAM5/MPI-OM) and some of them far from representing the observational mean (BCC-CM1, BCCR-BCM2.0, GISS-ER). Moreover, very similar values are obtained from the multiple runs for each model. Indeed, there is general recognition that the GCMs variability can be greater than inter-run variability (Wilby and Harris, 2006; Hawkins and Sutton, 2009).

This suggests that the multiple runs add little information in terms of the underlying climatology.

FIGURE 2.1: Southern Africa overall mean annual precipitation (mm) from CRUTS3 (1951–2000).



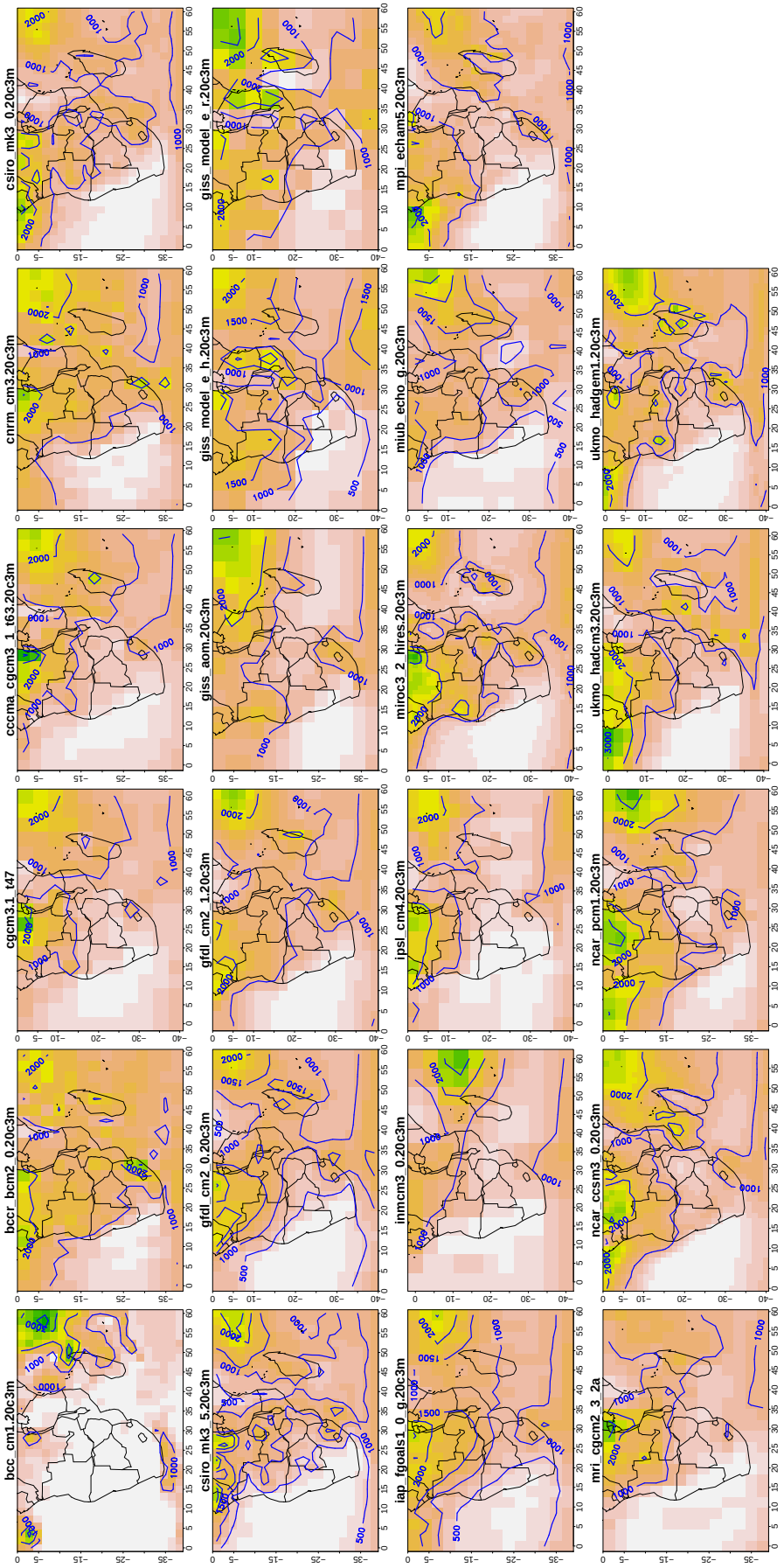
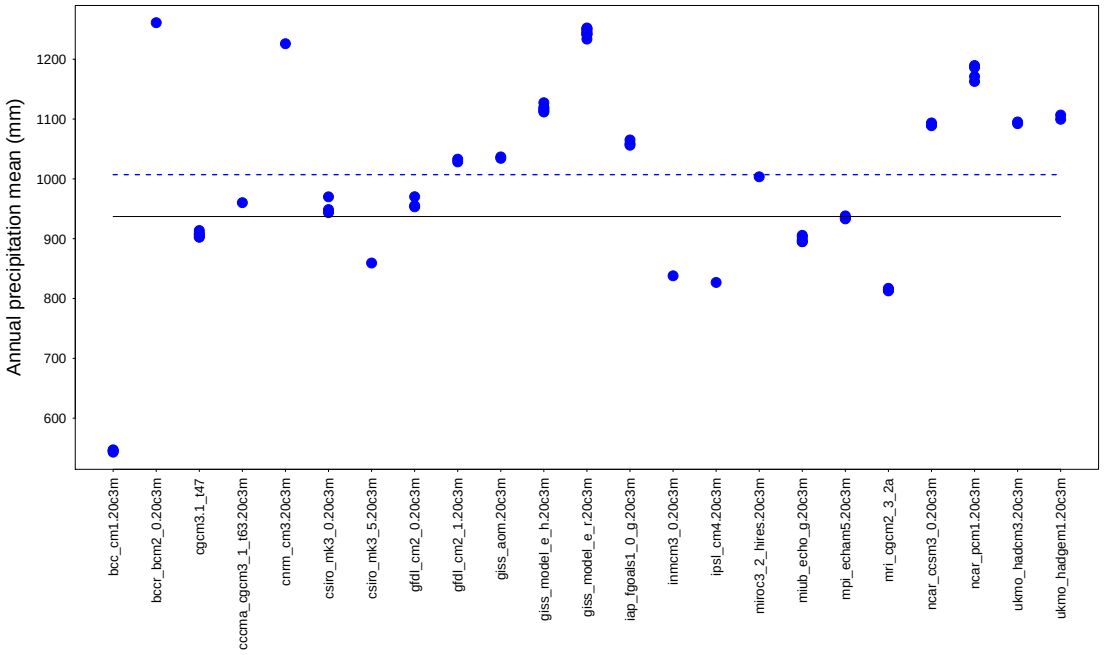


FIGURE 2.2: Southern Africa overall mean annual precipitation (mm) from 23 GCMs (1951–2000). Same colour scale as in Figure 2.1.

FIGURE 2.3: Overall annual precipitation averaged over the region as projected by the 23 GCMs and all available ensemble members (blue dots) for the period 1951-2000. The black solid line represents the CRUTS3 mean and the blue dashed line the mean of all model runs.



2.3.2 Monthly mean precipitation

To investigate the seasonal variation in precipitation, separate maps of mean precipitation were produced for each month of the year. For the studied region, as shown by the CRUTS3 dataset (see Figure 1.1), the wet season mainly corresponds to the austral summer, from October to April, with the largest rainfall amounts typically observed between December and March, while the dry season generally lasts from May to September. Only the southwestern Cape is an exception, with the wettest months occurring during the austral winter (May-August), due to its Mediterranean climate (see Section 1.2).

Although most of the GCMs tend to overestimate the monthly precipitation means around the subcontinent (images not shown here), the pattern of the seasonal precipitation is similar to the one observed. As already noticed in Section 2.3.1, some of the models produce higher values in the northwest area of the study region, especially during

the wet months. In general, the models do not reproduce the precipitation climatology well during the transition months. NCAR-CCSM3 and CSIRO-Mk3.0 overestimate the November monthly mean along the northwest-southeast gradient and UKMO-HadGEM1 overestimates the precipitation mean over the central region of the subcontinent during the wet season. The NCAR-CCSM3 also fails to reproduce the southern African rainfall pattern seen in the CRUTS3 plot, from January throughout March, but shows instead a semicircle feature over Zambia, southern Angola, Namibia and southeast South Africa.

2.3.3 Monthly standard deviation

Standard deviations were computed for each grid cell, separately for each month of the year, to investigate the variability in monthly precipitation amounts at each location (not shown here). The standard deviation is also used in the computation of the coefficient of variation (see Section 2.3.4), for which an approximately constant value is an assumption of the modelling framework (see Sections 2.3.4 and 3.3.1).

The pattern of standard deviations largely mirrors that of monthly means, with larger values during the wetter months (November, December, January and February) over the areas with higher monthly rainfall, such as the Indian Ocean around the Equator, the Atlantic Ocean along the Angolan coast and the northwest-southeast band over land.

During the dry season, especially from May to September, the standard deviation values are mainly around zero. Most of the GCMs also reflect this feature showing an overall agreement with the observational gridded data. The BCCR-BCM2.0, CCSM3, CGCM3.1, CNRM-CM3, CSIRO-Mk3.0 and 3.5, MIROC3.2 (hires) and UKMO-HadGEM1 better represent the mean rainfall standard deviation, as derived from CRUTS3, while only few, such as the MPI-ECHAM5, do not seem to capture the overall pattern.

In general the models tend to display greater variability of the summer precipitation compared with the climatology. This may reflect either a tendency to overestimate the standard deviation compared to the observational gridded dataset, or an artefact in the gridding procedure which will result in underestimation of variability.

2.3.4 Conditional coefficient of variation

The coefficient of variation (CV) is a normalized measure of dispersion of a probability distribution, defined as the ratio of the standard deviation to the mean. Here it is only computed for wet months with rainfall > 1 mm/month (from which the definition as 'conditional' on rain > 1 mm/month). The threshold of 1 mm/month has been chosen in order to reduce the variability in the calculated coefficients caused by persistently small monthly rainfall means. In a region where the estimated evapotranspiration losses are around 2–3 mm/day higher than precipitation (Shukla and Mintz, 1982), 1 mm/month has been considered irrelevant from hydrological and agricultural perspectives.

Throughout the wet season and over most of the subcontinent eighty percent of the conditional CV values fall between 0.2 and 0.9, with that range decreasing to 0.3-0.7 when only the wet season (Oct-Mar) is considered (see Figure 2.4).

Blank grid cells correspond to those locations/months where no rain above 1 mm is recorded/modelled. For some of the grid cells the coefficient of variation is lower than 0.5, though. In particular this occurs in a large area towards the north central part of the region. Since during the dry season and over certain areas within the study region the rainfall mean is close to 0 mm/month, the coefficient of variation is more sensitive to the value of the mean resulting in an extremely variable coefficient of variation. Indeed, the highest values, reaching values of 3.4 (greater than 10 if the threshold is fixed to 0 mm/month) and with an average of 2, are visible during the dry months (especially from May to August over the centre of southern Africa), and over the driest areas of the region (southwest coast and inland). However, for most of the GCMs, in the rest of the region the conditional coefficient of variation is in the range 0.5-1 (see Figure 2.5). Moreover, as expected from the GCMs results of the two previous sections, in some cases the values are slightly overestimated (see for example Figure 2.5, plot b).

The impact of assuming that the CV is roughly constant will be explored later in the development of models for monthly rainfall amounts (see Section 3.3.1).

FIGURE 2.4: Southern African conditional coefficient of variation from CRUTS3 (1951–2000).

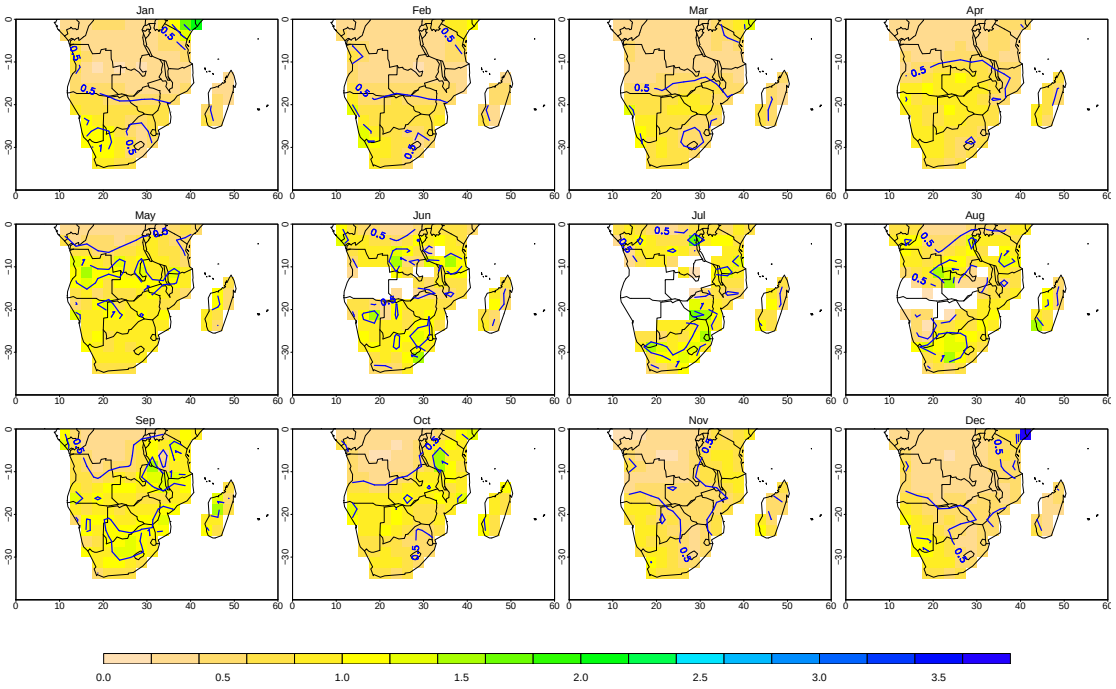
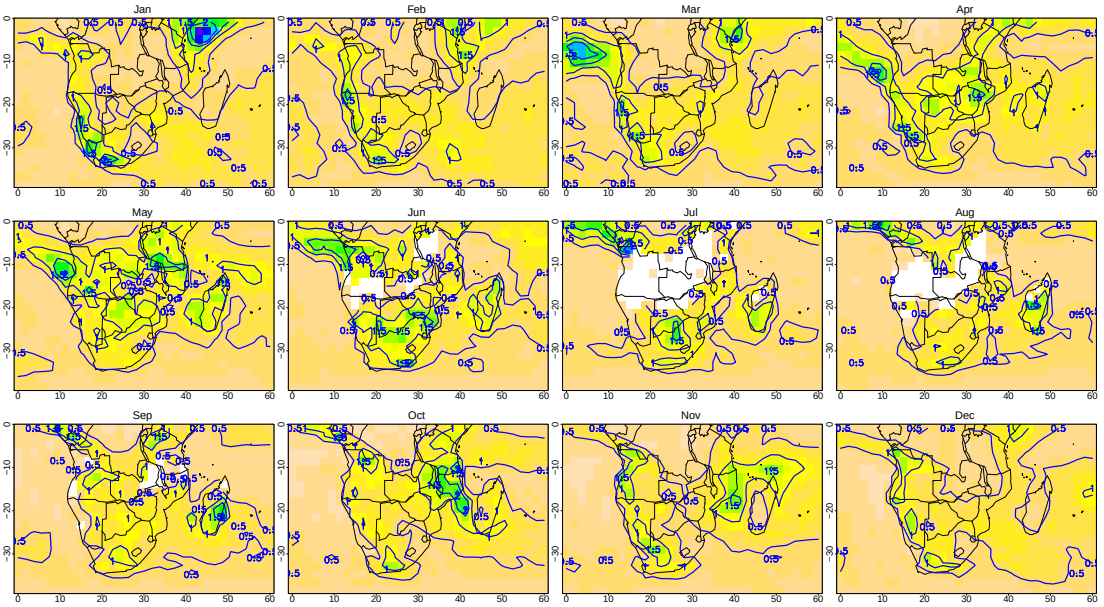
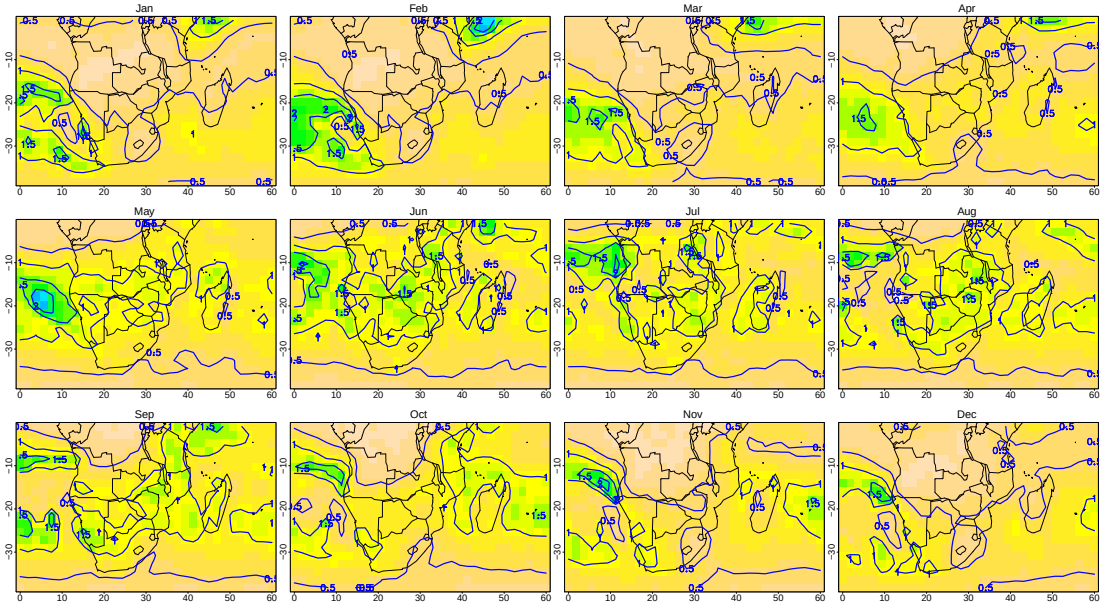


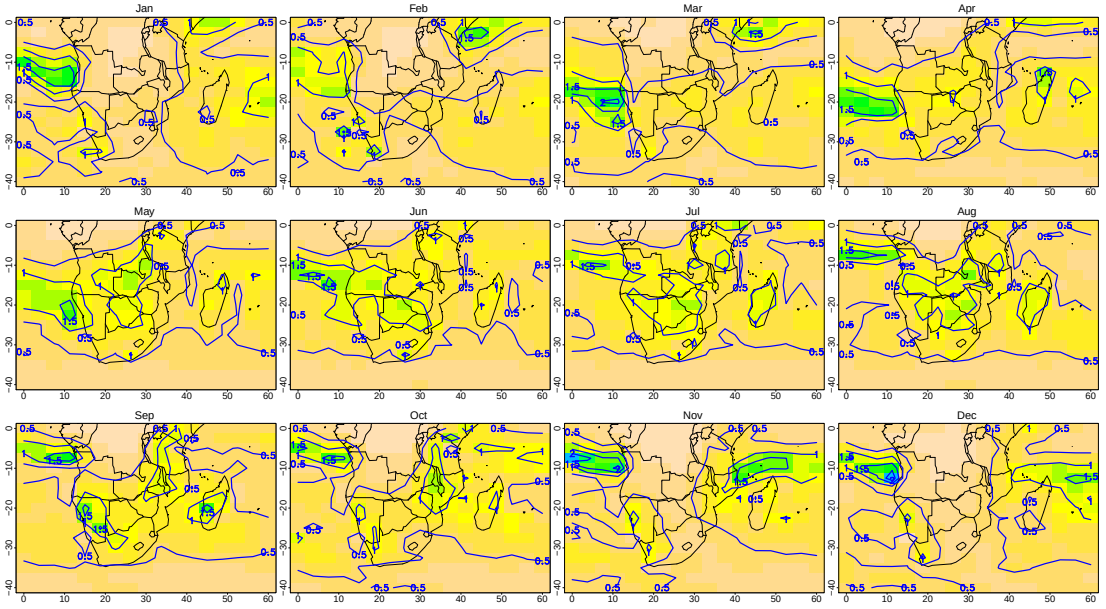
FIGURE 2.5: Southern African conditional coefficient of variation from 3 GCMs (1951–2000). Same colour range as in Figure 2.4.



(a) CSIRO-Mk3.5



(b) MPI-ECHAM5



(c) UKMO-HadCM3

2.3.5 Proportion of dry months

During the dry season, in areas of southern Africa individual months can experience zero, or very little, rainfall. To show this, for each month and grid node the proportion of months that the rainfall was below the fixed threshold of 1 mm/month was computed (see Figures 2.6 and 2.7). Later, such information will guide the modelling process where a two-stage approach is used to represent monthly rainfall occurrence and amounts separately. Only precipitation values of at least 1 mm will be used to fit the subsequent distribution of rainfall amounts.

As expected the highest proportions of dry months are concentrated during the austral winter and over the driest areas, such as the south-west coast of southern Africa. In qualitative terms the GCMs appear to reproduce the overall CRUTS3 seasonal pattern. However, models such as GISS-ER, GISS-EH, GISS-AOM, MIROC3.2 and UKMO-HadCM3 generated almost no or little dry months. On the contrary other climate models, such as CSIRO-Mk 3.5, projected higher than recorded proportions on the east of the region during winter months.

FIGURE 2.6: Southern African proportion of dry months from CRUTS3 (1951–2000).

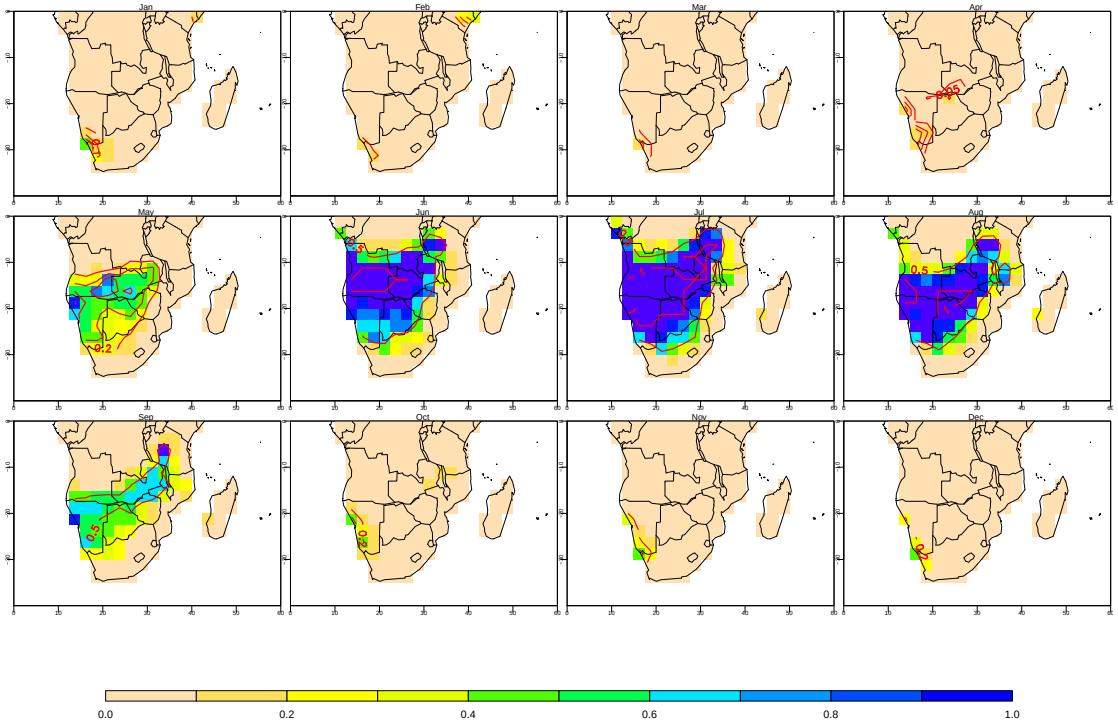
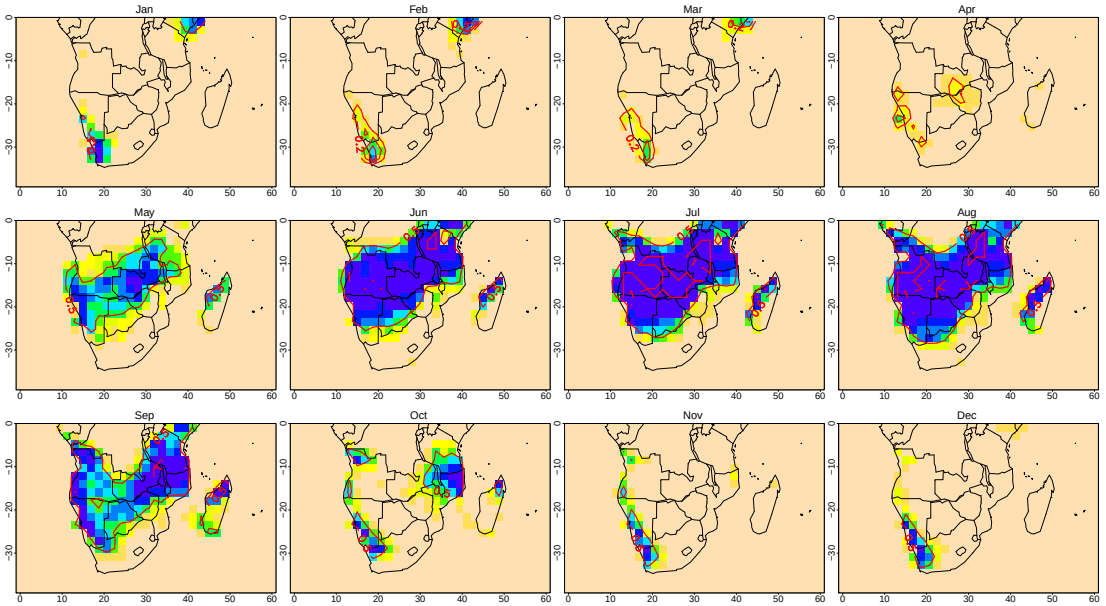
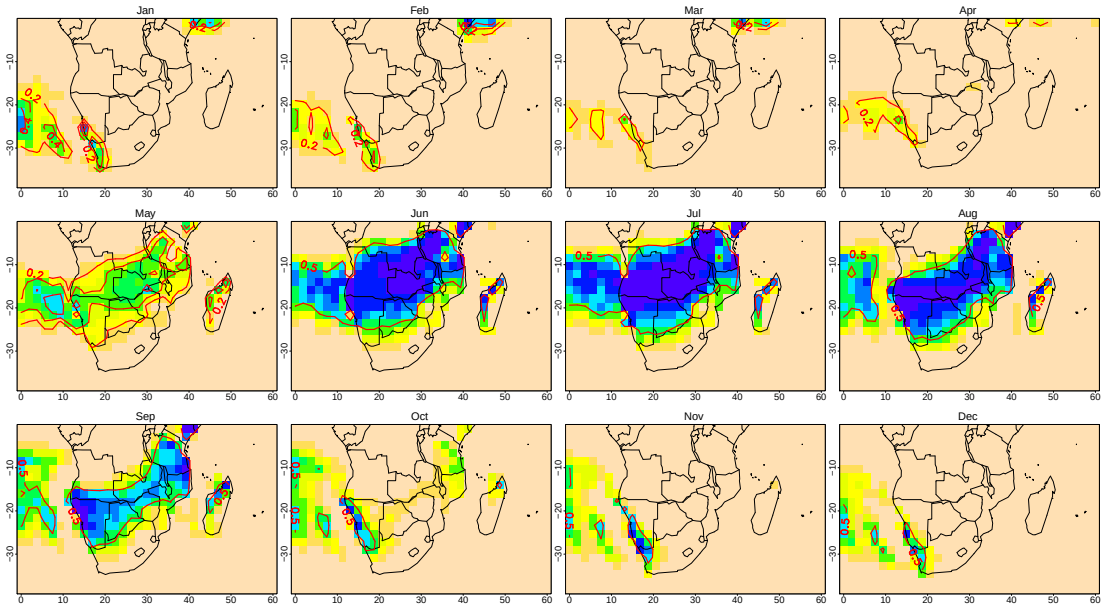


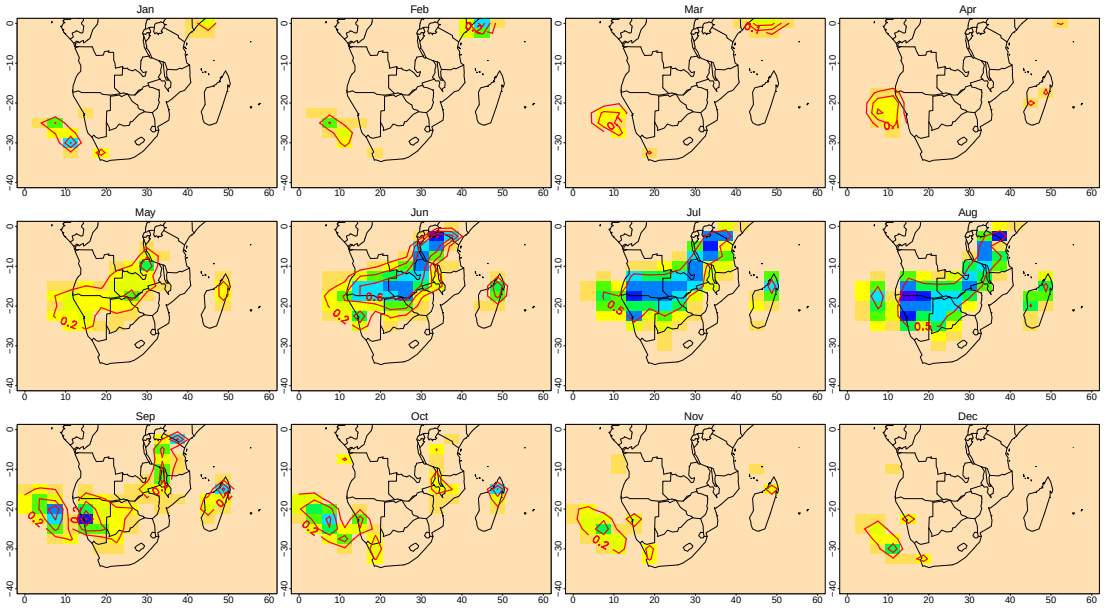
FIGURE 2.7: Southern African proportion of dry months from 3 GCMs (1951–2000). Same colour range as in Figure 2.6.



(a) CSIRO-Mk3.5



(b) MPI-ECHAM5



(c) UKMO-HadCM3

2.3.6 Amplitude and phase of the seasonal cycle

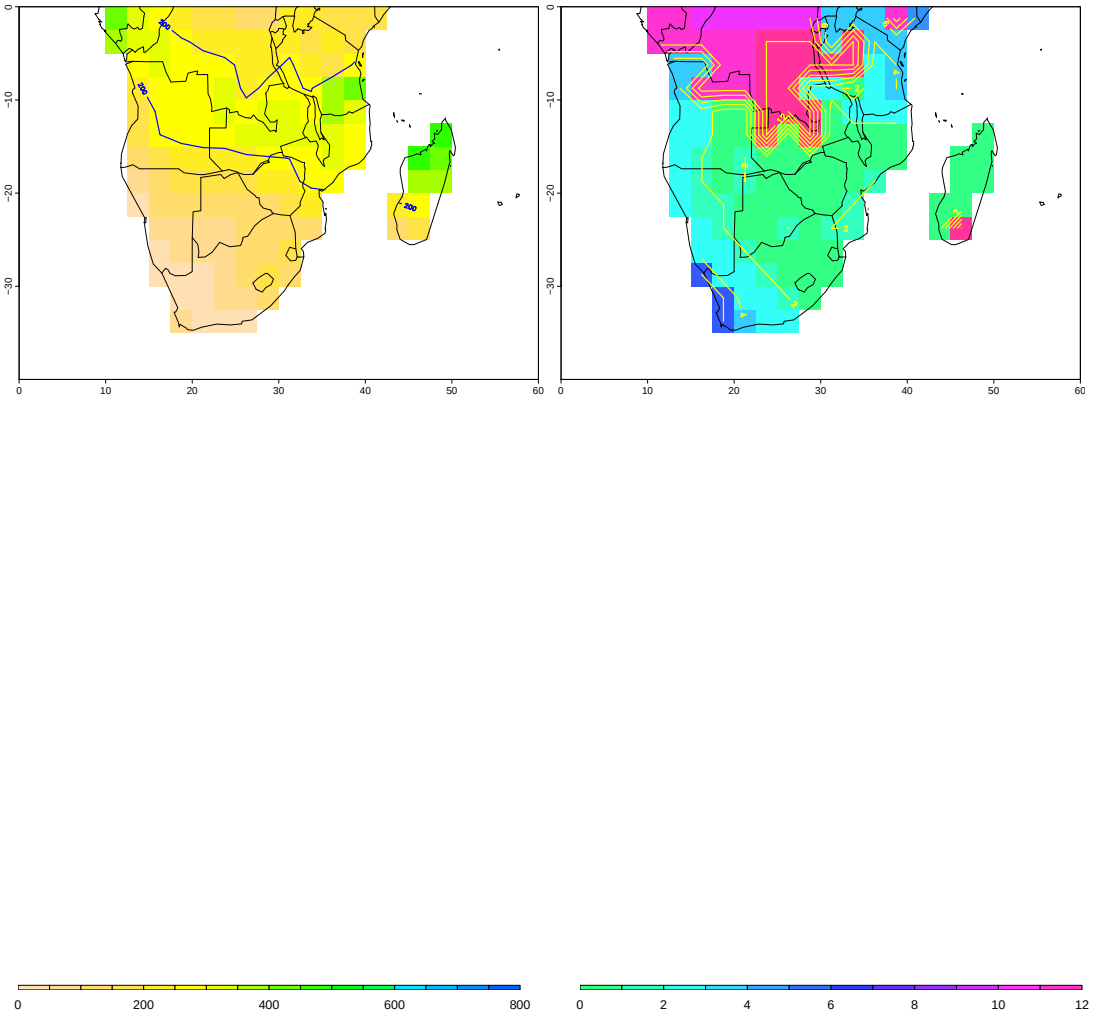
Besides the monthly rainfall mean, the amplitude and the phase of the seasonal cycle are used to assess the features of the seasonal cycle and the ability of the models to represent them. At each grid cell, the amplitude has been computed as the difference in mean precipitation between the wettest and driest months of the year.

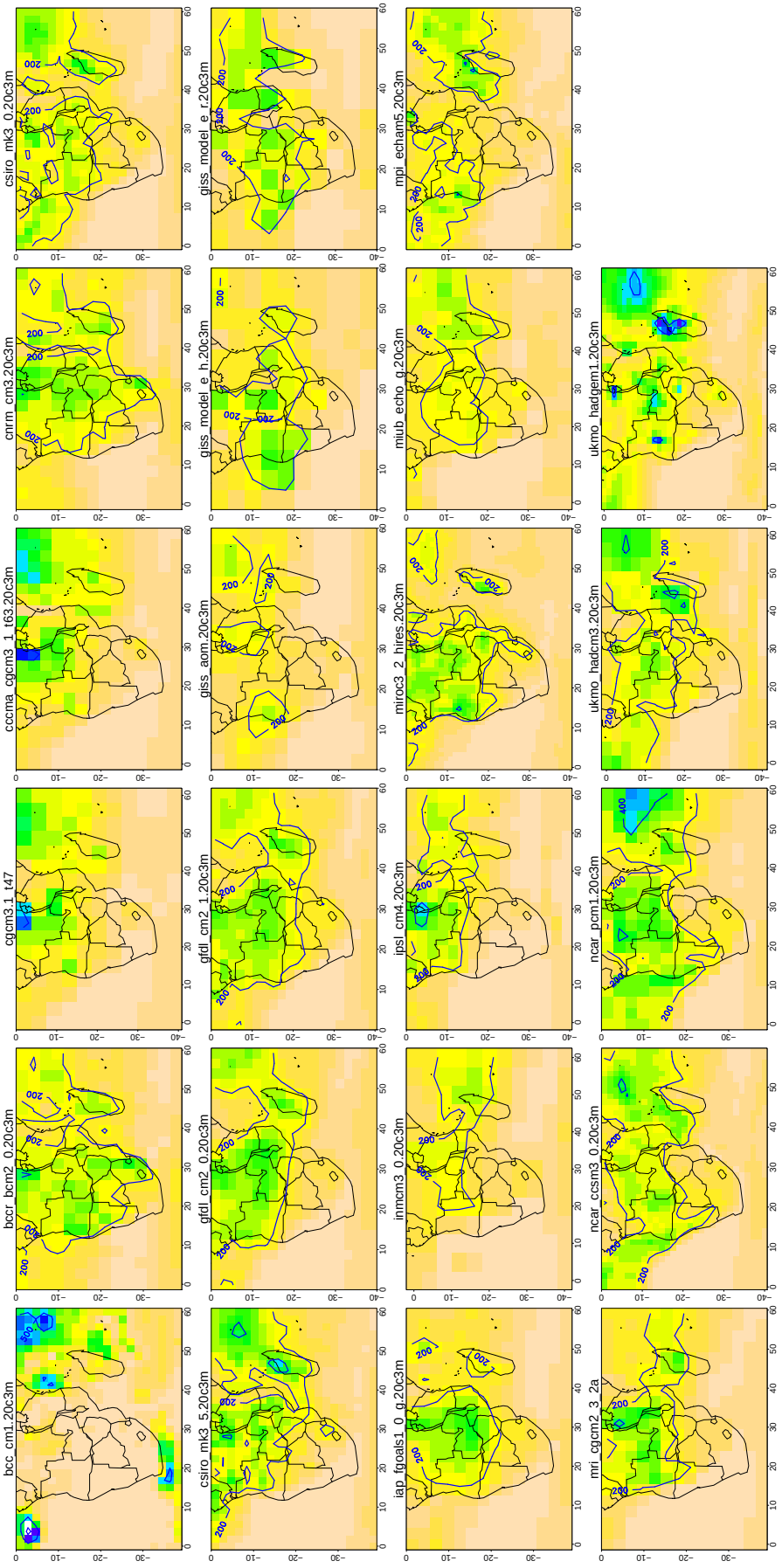
As expected, the peak of the seasonal cycle is generally found in the areas of higher mean rainfall (see Figure 2.8, left plot). For most of the GCMs the highest precipitation occurs in the same locations/months as the CRUTS3 dataset. However, the GCM amplitudes generally tend to be higher overall (see Figure 2.9, plot a)). Only for two models, the GISS-AOM and especially the BCC-CM1, is the structure noticeably different from that of the observations.

As a simple measure of the phase of the seasonal cycle, at each grid point, the month with the highest mean precipitation was recorded (see Figure 2.8, right plot and Figure 2.9, plot b)). The observational gridded dataset (Figure 2.8, right plot) clearly show four different features as far as this measure is concerned. The northern equatorial sector above the Congo basin receives maximum levels of precipitation during the earlier austral summer months (October – December). Eastern Africa experiences the peak in rainfall around March/April. The southwestern corner of Africa, over Cape Town, in contrast, experiences higher rainfall during the austral winter, in Jun, consistent with its Mediterranean climate. Over the remaining part of the region the peak of the seasonal cycle occurs in February/March.

The GCMs tend to reproduce some of the features, namely the October – December equatorial peak as well as the rainfall phase in the Mediterranean climate over the Cape. However, the phase of the seasonal cycle is not well reproduced in eastern Africa and the southeastern area of the subcontinent, where most of the models tend to place the peak in rainfall one or two month earlier.

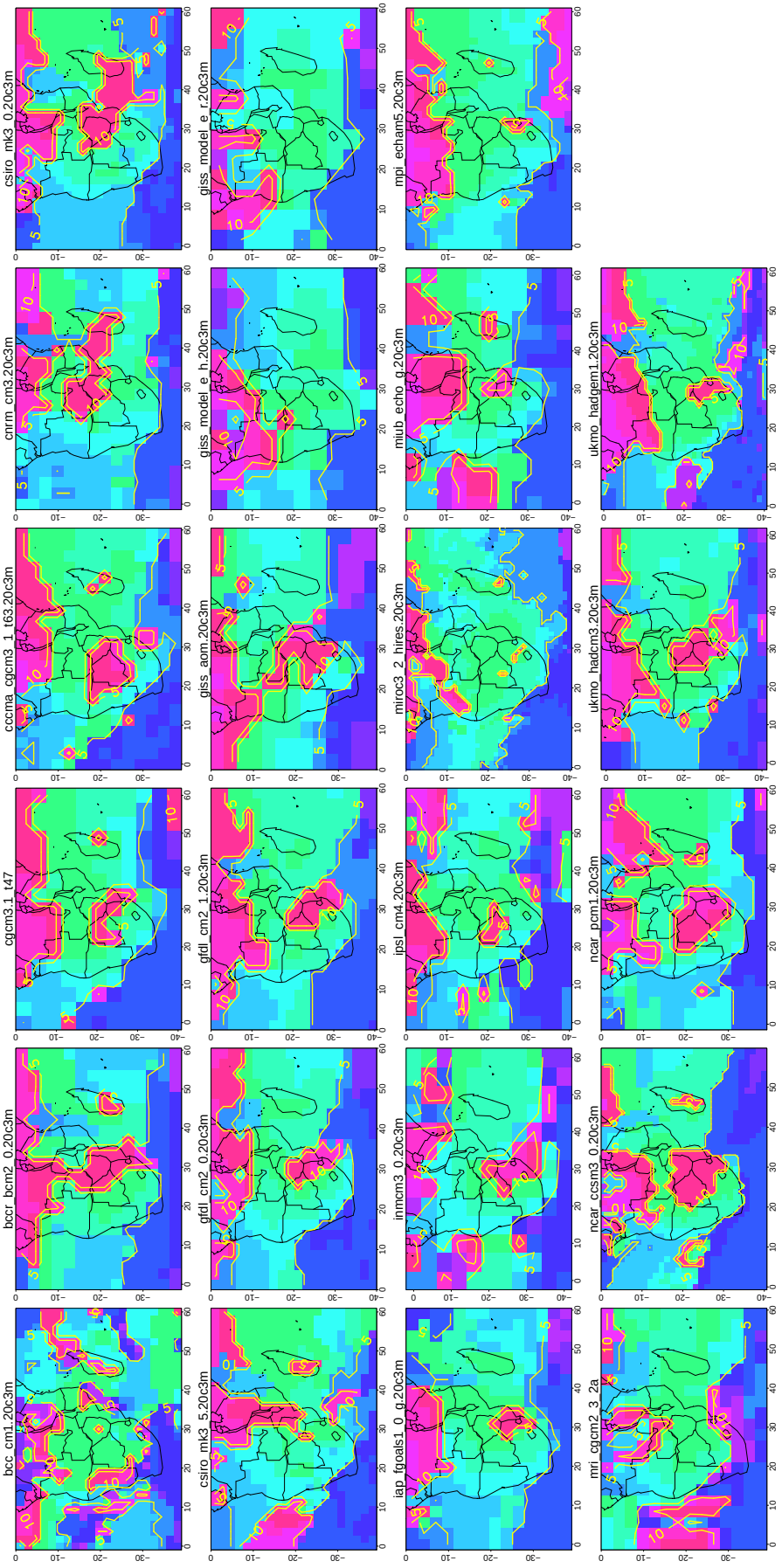
FIGURE 2.8: Southern African precipitation amplitude (mm) and phase (month of the year) of the seasonal cycle from CRUTS3 (1951–2000).





(a) Amplitude

FIGURE 2.9: Southern African precipitation amplitude and phase of the seasonal cycle from 23 GCMs (1951–2000). Same colour scales as in Figure 2.8.


csiro_mk3_0.20c3m
giss_model_e_r.20c3m
mpi_echam5.20c3m
cnrm_cm3.20c3m
giss_model_e_h.20c3m
miub_echo_g.20c3m
ukmo_hadgem1.20c3m
cccma_cgcm3_1_t63.20c3m
giss_aom.20c3m
miroc3_2_hires.20c3m
ukmo_hadcm3.20c3m
cgcm3.1_t47
gfdl_cm2_1.20c3m
ipsl_cm4.20c3m
ncar_pcm1.20c3m
bccr_bcm2_0.20c3m
gfdl_cm2_0.20c3m
inmcm3_0.20c3m
ncar_ccsm3_0.20c3m
bcc_cm1.20c3m
csiro_mk3_5.20c3m
iap_fgoals1_0_g.20c3m
mri_cgcm2_3_2a

(b) Phase

2.3.7 Annual trends

In order to investigate systematic changes in regional rainfall patterns, trends have been computed at each grid cell. To show the climatological trend, annual rainfall values were used instead of the monthly means. The annual values are aggregated over periods of twelve months, each month from July to June to reflect the hydrological year. The trends were computed by fitting straight line regressions through the series of annual means, and extracting the slope of the linear fit. Note that at this stage, the aim is not to establish whether the trends are statistically significant. The purpose is simply to produce a preliminary assessment of the quality of the GCM simulations and look at the kinds of structures present in the data that can later be added in the modelling framework.

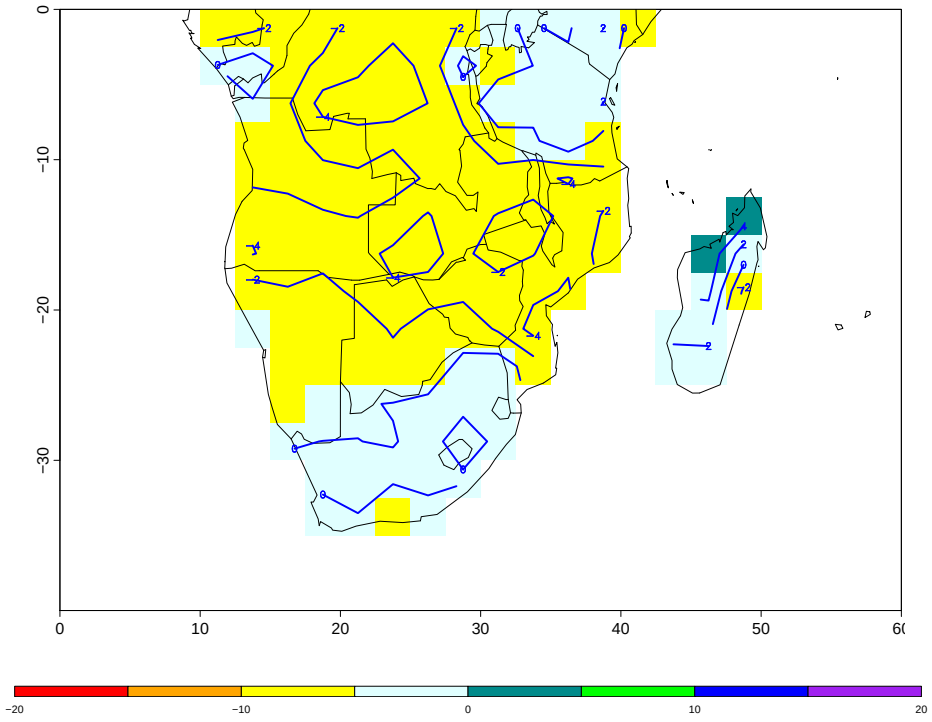
Figure 2.10 shows the results of this analysis for the CRUTS3 dataset. There is generally no or a slightly negative trend over the considered period over South Africa, Tanzania, southern Kenya, the Congolese coast and Madagascar. On the other end in Madagascar the trend appears to be positive and up to as much as 5 mm/50 years. However, the latter could be due to the poor station coverage used in the gridded product interpolation. Indeed, the found increase in rainfall has not been supported by studies investigating recorded precipitation trends in the island (Morishima and Akasaka, 2010; Vincent et al., 2011).

The remaining part of the study area shows a drying trend, reaching about -5 mm/50 years at the most, with more pronounced drying spots in Zambia, western and coastal Mozambique, northern Zimbabwe, central Democratic Republic of Congo and the Namibian coast. In general no strong signal is noticeable from the plot.

Figure 2.11 shows the corresponding analysis for each of the 23 GCMs. A number of the models reflect, in turn, some of the observed features such as wet trend over part of Tanzania (CGCM3.1, NCAR-PCM1, ECHO-G), Madagascar (CGCM3.1, FGOALS-g1.0, INM-CM3.0,) and South Africa (CNRM-CM3, GFDL-CM2, GISS-EH, FGOALS-g1.0, NCAR-PCM1 and UKMO-HadGEM1). However, in general there is little agreement among the models in terms of trends over the region. A group of GCMs, including CGCM3.1, FGOALS-g1.0, IPSL-CM4, GISS-EH, MRI-CGCM2.3 and UKMO-HadGEM1, shows a general increasing trend in annual precipitation over the land, whilst

others (GISS-AOM, GISS-ER, NCAR-CCSM3.0, NCAR-PCM1, and UKMO-HadCM3) show drying trends in different locations than these in the observational data. Some of them show a widespread drying trend along the coastal areas of Angola and Namibia (BCCR-BCM2.0, CGCM3.1, CNRM-CM3 and FGOALS-g1.0).

FIGURE 2.10: Southern African overall annual trend (mm/50 years) from CRUTS3 (1951–2000).



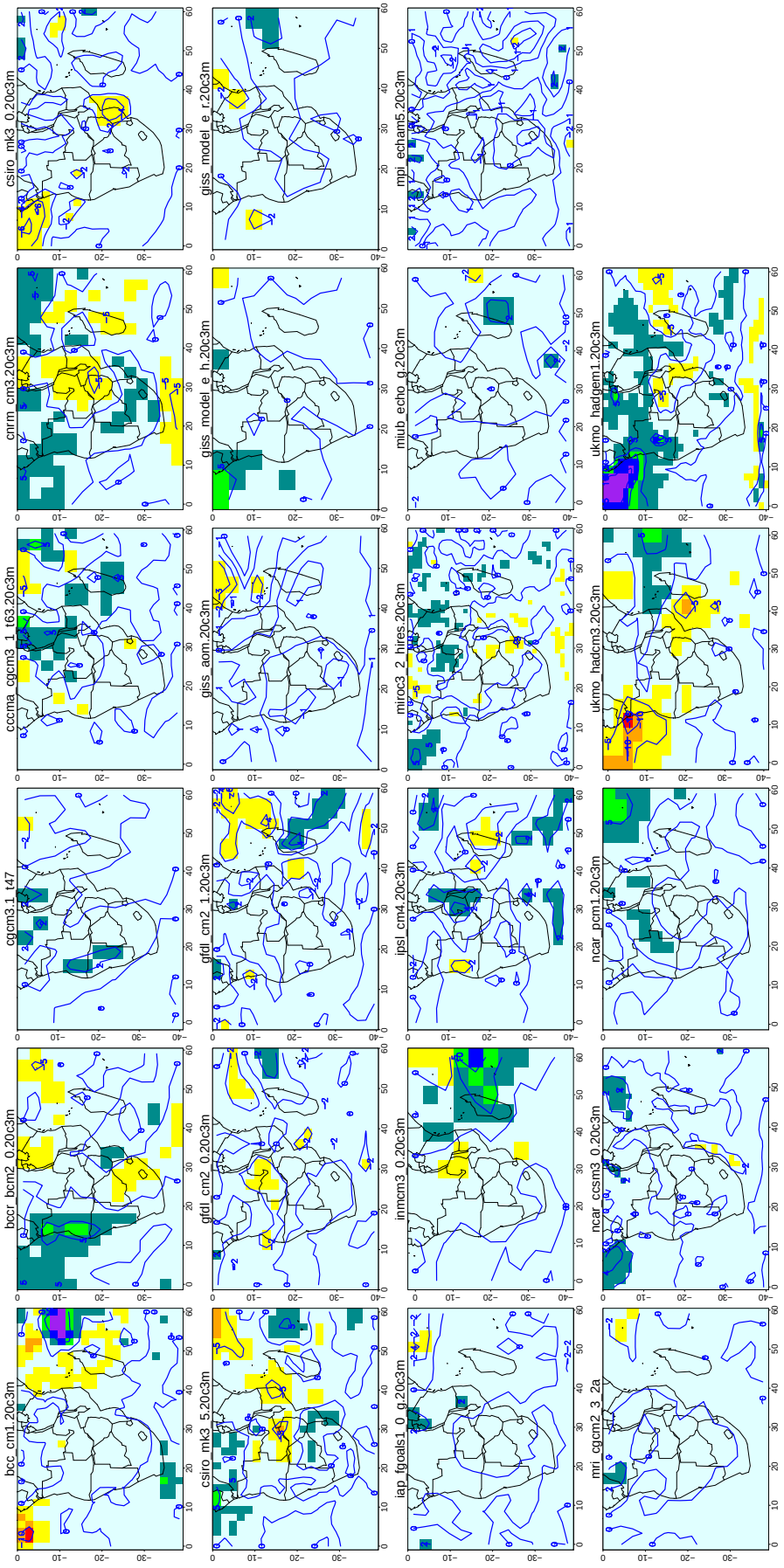
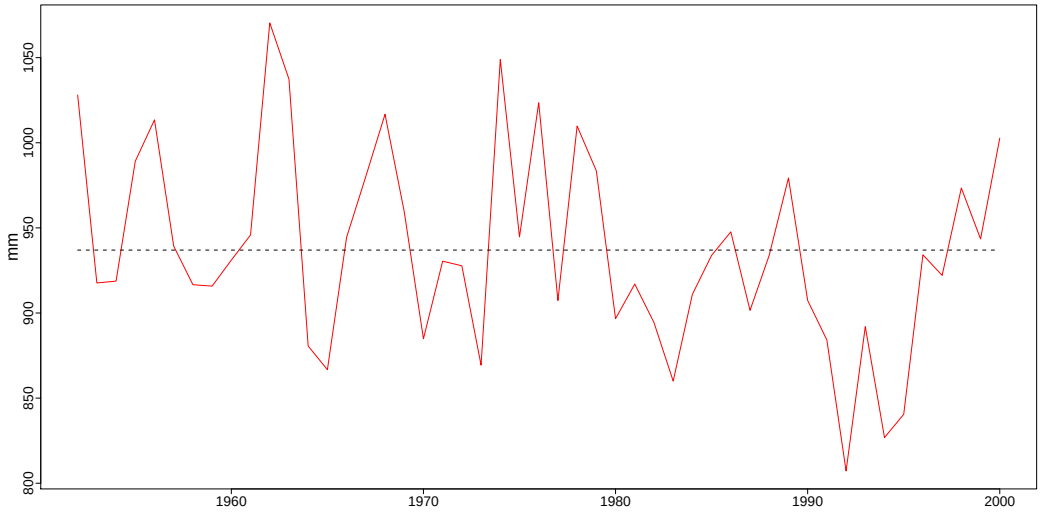


FIGURE 2.11: Southern African overall annual trend (mm/50 years) from 23 GCMs (1951–2000). Same colour scale as in Figure 2.10.

To put the linear trends in context, Figure 2.12 shows the time series of regionally averaged annual precipitation means. As might be expected from the linear trend results, there is little clear signal of overall change here although interannual variability is strong (which contributed to the well documented severe droughts of the 1980s and early 1990s) and visible in the figure below. Such results are in agreement with Mason and Jury (1997) and Richard et al. (2000).

FIGURE 2.12: Overall annual precipitation mean averaged over the region from CRUTS3 (1951–2000). The black line represents the series mean.



2.4 Conclusion

The results of the preliminary analysis show clearly that the rainfall climate of southern Africa is characterised by a pronounced seasonal cycle, with a dominant northwest/south-east convection system during the summer season and Mediterranean climate over the southwestern Cape. This pattern leads to a rainy austral summer over almost the whole region, whereas in winter the subcontinent is characterised by spatially and temporally widespread dry periods. During these months (mainly May to September) the region is associated with stable dry conditions.

Some of the GCMs have been shown to represent the main patterns of the latter half of the 20th century rainfall in terms of overall mean and seasonal cycle in qualitative

but not necessarily in quantitative terms. In particular, the linear trend is poorly represented with respect to the observational gridded data and no consistent pattern is identifiable among the GCMs (see Figure 2.11).

In summary, this chapter has investigated the rainfall characteristics of the region and has examined the ability of the climate models to reproduce the observed values and assess the potential use for impact study applications. After having identified weaknesses and abilities of the climate models, the modelling exercise was started with the choice of avoiding using climate model's rainfall and simulating it instead with a statistical model described in the next chapters.

Chapter 3

Southern African monthly rainfall variability

3.1 Introduction

In the previous chapter it was established that the GCMs representation of precipitation in the study region can be poor. On way to avoid some of the potential problems associated with this is to build statistical models to link the observed precipitation with larger-scale structures that may be better represented. Therefore relationships between southern African rainfall and large-scale climate factors have been investigated, building upon the current knowledge of the structures controlling subcontinental rainfall variability (Section 1.2.1).

The analysis presented here is based on generalized linear models (GLMs). The GLMs are fitted to 20^{th} century observational data to characterise the dependence of monthly precipitation upon the climate indicators of interest. In contrast with many of the correlation-based analyses that have previously been used to investigate controls on precipitation in the region, GLMs allow the investigation of the relationships between different components of the climate system (geographical and climatic drivers) simultaneously (Chandler, 2005).

As well as characterising the effect of the covariates upon precipitation, because GLMs

are explicitly probabilistic they provide the opportunity to simulate multiple precipitation sequences that are consistent with any given set of atmospheric drivers while recognising that the dependence is not deterministic. Thus, assuming that the covariate-precipitation relationship remains unaltered in the future (one of the assumptions of statistical downscaling, see Section 1.5.1), the methodology allows information from GCM future projections to be used to drive GLM simulations of future precipitation. This can be done without the direct use of GCM precipitation outputs, which is helpful given the lack of skill in the current GCM simulations of hydrological processes. Hence, the identified relationships can be used for the downscaling of GCM outputs, bridging the gap between the coarse resolution of those climate models and the required resolution for impact assessment studies (Hewitson and Crane, 1996; Wilby and Wigley, 1997; Maraun et al., 2010) (see application in Chapter 5).

3.2 Area considered and Data used

An initial modelling exercise aimed to fit GLMs to data from the entire region for which exploratory analyses were reported in Chapter 2. However, diagnostics revealed that the model performance was poor in the equatorial region. There could be two reasons for this: firstly, station data are sparse in the equatorial region so that the quality of the gridded data here is questionable; and secondly, the tropical climate of this region is so very different from that of the rest of the southern Africa that it is very difficult to represent the entire subcontinent adequately using a single statistical model. Fortunately, for the final purposes of studying implication of water resources for agriculture, the equatorial region is of little interest; it has therefore been excluded from the modelling exercise resulting in a final area of study between 12° – 40° S and 0° – 42° E. Because of data availability the study focused on the second half of the 20^{th} century. Specifically 50 years, from 1957 to 2006, of contemporary records were used in the study. Twentieth century precipitation data are derived from the CRUTS3 gridded product (CRU, 2008, see Chapter 2). Gridded datasets are widely used in climate research since they provide values of the variable of interest at any spatial location, and are especially useful for those areas characterised by a low and sparse density of meteorological stations, such as the African continent (Washington et al., 2006). Moreover, they are available for a

relative long period of time. In particular, a number of previously published papers used the same kind of dataset to study the African climate (Jury et al., 2004; Conway et al., 2009; Shongwe et al., 2009), justifying its use in the work presented here.

Even though many other studies have used gridded precipitation data, it should be recognized that the gridding procedure may introduce artefacts due to the locations of available stations and the chosen interpolation method. However, since those artefact typically operate at a fairly localised scale these gridded data were considered, with one exception, suitable for characterising the large-scale regional structures and relationships. The exception relates to the grid cells centred over Angola, for which no station data were available after the year 1972. Therefore data from these grid cells, from 1973 onwards, have been excluded from the modelling process.

Table 3.1 summarises all the climate variables investigated as potential drivers for southern African precipitation, the data sources and the data references. All time series are monthly and reflect the dominant indices identified in the literature (Section 1.2.1).

An initial set of indices included atmospheric and oceanic fields that are physically meaningful and affect southern African rainfall variability, such as sea surface temperatures, sea level pressures and relative humidity. For some climate phenomena and teleconnections involving these fields, existing indices were considered (e.g. Nino3.4, NAO, SAM; see Table 3.1). In order to precisely select the most appropriate indices, deeper analyses were also performed. The aim was not only to understand the behaviour of the potential drivers, but also to identify possible patterns that are highly correlated with southern African rainfall. To do so principal component analysis (PCA) and canonical correlation analysis (CCA) were used.

TABLE 3.1: Climate drivers of southern African rainfall variability considered in the modelling process.

Index	*Reference*	*Data Source*	*Description*
El Niño Index (NINO)	Rayner et al. 2003	Met Office Hadley Centre	Normalized SSTs for Nino3.4 region
Southern Oscillation Index (SOI)	Ropelewski and Jones 1987	Climate Research Unit-UEA	Normalized pressure difference between Tahiti and Darwin
Quasi-Biennial Oscillation (QBO)	Calculated at PSD from NCEP/NCAR Reanalysis	NOAA/ESRL Physical Sciences Division	Zonal average of the 30mb zonal wind at the Equator
North Atlantic Oscillation (NOA)	Jones et al. 1997	Climate Research Unit-UEA	Normalized pressure difference between a station on the Azores and one on Iceland
Indian Ocean Dipole (IOD)	Saji et al. 1999	Met Office Hadley Centre	Anomalous SST gradient between the western equatorial Indian Ocean and the south eastern equatorial Indian Ocean
Indian Ocean SSTs	Smith et al. 2008	NOAA/National Climatic Data Center	Normalized SST over the Indian Ocean (40°E – 11°E, -20°N – 20°N)
Southern Hemisphere Temperature (SHT)	Brohan et al. 2006	Met Office Hadley Centre	Standardised Southern Hemisphere Surface Temperature
Southern Annular Mode (SAM)	Marshall 2003	Station Data	Mean sea level pressure difference between 40° and 65°S
South West Indian Ocean Dipole (SWIO)	Smith et al. 2008	NOAA/National Climatic Data Center	Extended Reconstruction Sea Surface Temperature Differences between north-west and south-west Indian Ocean
Brandon Marion Index (BMI)	Kalnay et al. 1996	NOAA/ESRL Physical Sciences Division	Principal component of South Indian Ocean Pressure representing differences between St.Brandon and Marion Islands
Agulhas SSTs	Smith et al. 2008	NOAA/National Climatic Data Center	Leading principal component for Sea Surface Temperature in the Agulhas region
Benguela SSTs	Smith et al. 2008	NOAA/National Climatic Data Center	Leading principal component for Sea Surface Temperature in the Benguela region

Index	*Reference*	*Data Source*	*Description*
South Atlantic SSTs	Smith et al. 2008	NOAA/National Climatic Data Center	Leading principal component for Sea Surface Temperature in the South Atlantic Ocean
Southern African Relative Humidity (RHUM)	Kalnay et al. 1996	NOAA/ESRL Physical Sciences Division	Leading principal component of Relative humidity over southern Africa (12.5° – 42.0°E, -35.0°– 0.0°N)
Pacific Decadal Oscillation (PDO)	Zhang et al. 1997	Met Office Hadley Centre	Leading principal component of North Pacific monthly sea surface temperature variability
Atlantic Multidecadal Oscillation (AMO)	Van Oldenborgh et al. 2009	Met Office Hadley Centre	Sea Surface temperatures in the northern Atlantic
Jet Stream Intensity and Position	Kalnay et al. 1996	NOAA/ESRL Physical Sciences Division	Intensity and Position of jet stream at 200mb

The derived indices were produced as standardised anomalies with respect to the 1961 – 1990 climatology. That period was chosen because commonly used as control period by the United Nation WMO (World Meteorological Organization) and in climate science (Hulme, 1992; Hulme et al., 1995).

3.2.1 Relative humidity dataset

The 850 hPa relative humidity data used here are from the NOAA NCEP reanalysis product (Kalnay et al., 1996). Although homogeneous techniques are used to generate the reanalysis data, the effect of changes in the atmospheric observing system is unavoidable. One prominent impact has been the introduction, from 1979 to present, of satellite data, particularly over the southern hemisphere, causing inhomogeneities in the form of step changes between pre and post-1979 reanalysis data (Kistler et al., 2001; Sterl, 2004; Tennant, 2004). To remove the effects both of these inhomogeneities and of seasonality, we used additive models (Bowman and Azzalini, 1997). At each grid cell, a model of the form:

$$\mathrm{RHUM}_t = a + s(t) + [b \times I\,(\mathrm{Year} < 1979)] + \left[c \times \cos\left(\frac{2\pi \times \mathrm{Month}}{12}\right)\right] +$$

$$\left[d \times \sin\left(\frac{2\pi \times \text{Month}}{12}\right)\right] + \varepsilon_t$$

was fitted to the original monthly time series. Here, the trend term $s(t)$ is a smooth function of time, represented nonparametrically so as to avoid imposing artificial structures that may be unrealistic; the cosine and sine terms represent the seasonal cycle; and $I\,(\text{Year} < 1979)]$ is an indicator variable taking the value 1 for all observations prior to 1979 and 0 thereafter. Estimation of the smooth trend $s(t)$, along with the coefficients a, b, c and d, was carried out using the `gam()` routine (Hastie, 2011). The coefficient b is the magnitude of the step associated with the introduction of satellite data. A deseasonalised and destepped series can thus be defined as

$$\text{RHUM}_t^* = a + s(t) + \varepsilon_t \ ;$$

this is referred to as 'adjusted relative humidity'. The process is illustrated for a single grid cell in Figure 3.1.

3.2.2 Wind speed dataset

The effect of the jet stream upon precipitation in a particular location can be expected to depend both upon the jet stream intensity and upon its position relative to the location of interest. 200 mbar wind speed reanalysis data were used to define the intensity and location of the jet stream, corrected for inhomogeneities in the same way as for the relative humidity field. At a particular time t, the jet stream position (POS_t) was defined to be the latitude corresponding to the maximum zonally-averaged winds over the region $(60°\text{S} - 20°\text{S}) \times (0°\text{E} - 45°\text{E})$. The corresponding zonally-averaged maximum intensity was then extracted. The resulting time series was standardised to form an index (INT_t say) of jet stream intensity.

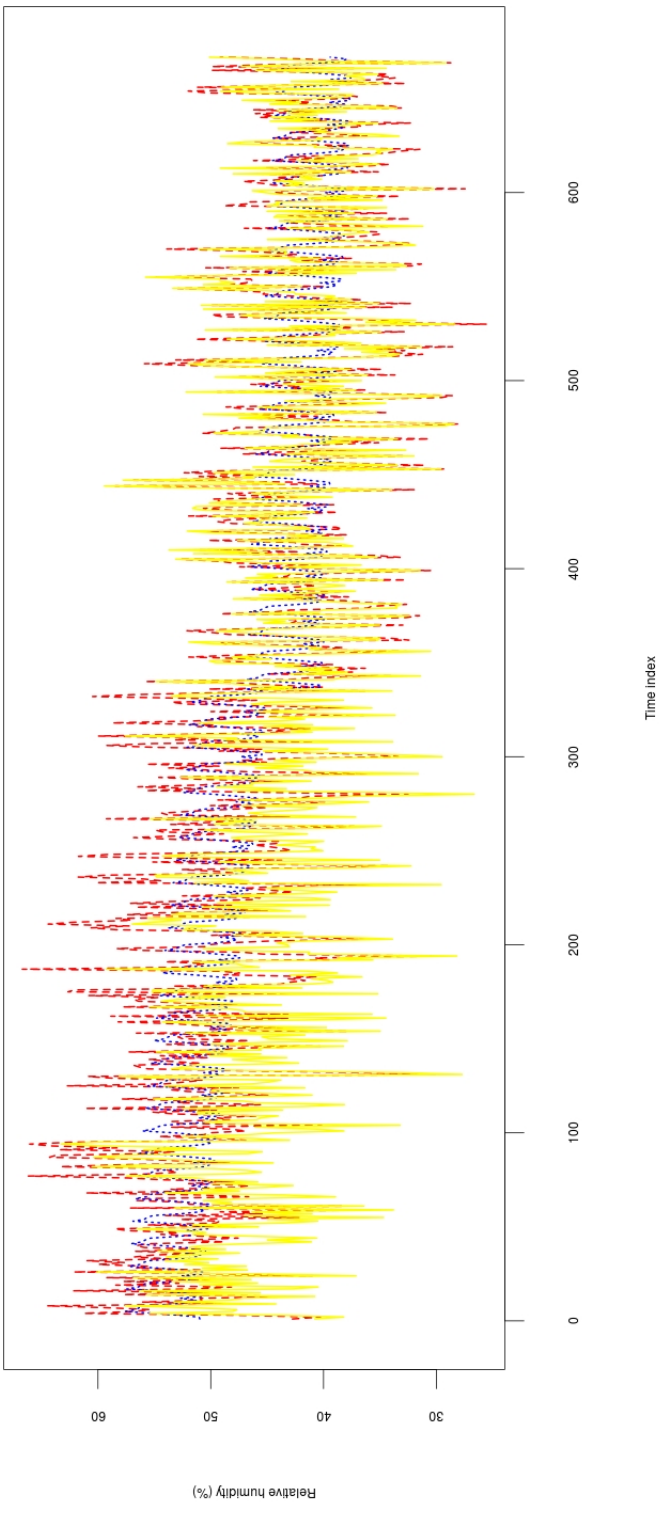


FIGURE 3.1: Illustration of adjustment to relative humidity time series from a single grid cell. Dashed red line: original series RHUM_t. Dashed blue line: estimate of $a + s(t) + [b \times I\,(\text{Year} < 1979)]$ in fitted GAM. Solid yellow line: adjusted series RHUM_t^*.

3.3 Modelling framework

3.3.1 Generalized linear models for rainfall

To explore the simultaneous effect of the many factors suggested as potential controls upon precipitation in southern Africa, analysis was based on GLMs, which can be regarded as an extension of multiple regression techniques. GLMs have been used widely in statistical science and recently explored in some papers in climate research (e.g. Buishand et al., 2004; Chandler, 2005; Furrer and Katz, 2007; Fealy and Sweeney, 2007). Chandler and Wheater (2002) and Yan et al. (2002) demonstrated the power of GLMs for analysing and understanding relationships among components of the climate system and in quantifying changes in it. The flexibility of GLMs, coupled with the fact that they are easily interpretable and computationally inexpensive, supports their application here. Here a GLM for monthly gridded precipitation data is defined by setting up a probability distribution for each time point and grid cell, from which the corresponding observation is considered to be drawn. These individual distributions are themselves constructed from the values of other variables (referred to as covariates) specific to that time point and grid cell.

In the GLM framework, the observations are regarded as realised values of a vector of random variables $\mathbf{Y} = (Y_1, \ldots, Y_n)'$ say, all assumed to be generated from the same family of distributions (e.g. Poisson, normal, gamma). Within this family, the expected value of Y_i is modelled as:

$$g(\mu_i) = \mathbf{x_i'}\beta = \eta_i$$

for some function g(.) (the link function), given a vector of $\mathbf{x_i'}$ of covariates and vector of coefficients β.

The GLM is then composed of three elements: a choice of distribution (e.g. binomial or gamma); a linear predictor $\eta = \mathbf{x}\beta$; a link function g such that $E(Y) = \mu = g^{-1}(\eta)$.

Following Coe and Stern (1982), a two-stage approach has been used here to model rainfall occurrence (> 1 mm/month rainfall) and amount separately. The given threshold of 1 mm rather than 0 mm to differentiate a wet from a dry month avoids some of the ambiguities that can be associated with the recording or estimation of very small rainfall amounts, and is practically insignificant for agricultural and hydrological purposes.

The rainfall occurrence (pattern of dry/wet months) has been described using logistic regression:

$$\ln \frac{p_i}{1-p_i} = \mathbf{x_i}'\beta \tag{3.1}$$

where p_i is the probability of rain for the i^{th} case in the data set conditional on a covariate vector $\mathbf{x}_i$ with coefficient vector β.

To model rainfall amounts during wet months, gamma distributions have been used. The rainfall amount for i^{th} wet month has, conditional on a covariate vector ξ_i and coefficient vector γ, a gamma distribution with mean μ_i and shape parameter ν, where:

$$\ln(\mu_i) = \xi_i'\gamma \tag{3.2}$$

The shape parameter ν is assumed constant for all observations; this corresponds to a constant coefficient of variation. For the purpose of assessing which climate covariates have an effect upon precipitation variability, the assumption of a constant CV is not considered acceptable. Particularly when using a robust estimate of the standard error (Chandler and Bate, 2007), able to cope with misspecification in the variance structure (see Section 3.3.3).

The gamma distribution has been widely used to represent monthly precipitation. It has been shown to be a good choice for water resources application in developing countries, because it is flexible enough to represent a variety of distribution shapes and rainfall regimes (Husak et al., 2007). In addition, GLM provides a well understood and computationally efficient statistical framework for the fitting procedure. Therefore, for the purpose of identifying relationships between rainfall and potential climate drivers, where the precise distribution is not critical, the use of a gamma distribution has been considered satisfactory.

After choosing an appropriate set of covariates, the parameter vectors β and γ were estimated using maximum likelihood using the `glm()` routine in R. This assumes that the monthly rainfalls are conditionally independent given the covariates, whereas this is unlikely to be the case. The analysis therefore needs to account for temporal and spatial dependence as discussed later.

Interactions, representing the possibility that the effect of one covariate depends on the

values of others, were also considered. This allowed the models to represent, for example, differences in the phase of seasonal cycle around the large area of study, as well as seasonal and regional variation in the strength of dependence between successive wet months, and in the effects of any climate covariate. Those interactions are represented by adding into the models covariates whose value is the product of the interacting predictors (Chandler and Wheater, 2002). In particular, following Chandler and Scott (2011) if a linear predictor η_i depends on two covariates (say x_{1i} and x_{2i}), where the effect of one depends on the value of the other, one might plausibly represent this dependence by using:

$$\eta_i = \beta_0 + \beta_1 x_{1i} + \beta_{2i} x_{2i}$$

$$\text{with} \quad \beta_{2i} = \gamma_0 + \gamma_1 x_{1i}$$

therefore, replacing in the first equation:

$$\eta_i = \beta_0 + \beta_1 x_{1i} + (\gamma_0 + \gamma_1 x_{1i}) x_{2i}$$

$$\eta_i = \beta_0 + \beta_1 x_{1i} + \gamma_0 x_{2i} + \gamma_1 x_{1i} x_{2i} \tag{3.3}$$

Terms that are not interactions, and involve single covariates, are here referred to as main effects.

3.3.2 Model checking

To check for unexplained structure, mean Pearson residuals were computed for different subsets of observations, separately for the occurrence and amounts models. For an observation Y_i the Pearson residual is defined as follows (Chandler, 2005):

$$r_i = \frac{(Y_i - \mu_i)}{\sigma_i}$$

where Y_i is the observed response for the i^{th} case, μ_i the modelled mean and σ_i the modelled standard deviation. If the fitted model is correct the Pearson residuals should all come from distributions with mean zero and variance 1; hence the standardised mean

from any subset of n independent observations: $\bar{r}\sqrt{n}$ where $\bar{r} = \frac{\Sigma r_i}{n}$, say, should itself have mean zero and variance 1. In addition, if n is large then the standardised mean will have approximately a normal distribution. Thus, if standardised mean Pearson residuals are calculated for many different subsets (for example, for each grid cell separately for each month of the year) and if the model is adequate, then around 95% of them should lie between -2 and +2.

For the rainfall amounts model, a further useful measure of performance is given by the Anscombe residuals. They are defined in such a way as to be approximately normally distributed if the studied data follow the here assumed gamma distribution. These residuals are used to check the distribution assumption of the models and for the gamma distribution the Anscombe residual takes the form:

$$r_i = \left(\frac{Y_i}{\mu_i}\right)^{1/3} \tag{3.4}$$

A quantile to quantile plot of the Anscombe residuals against a normal distribution provides an easy test and a resulting straight line implies a confirmation that the gamma assumption provides an adequate representation (Chandler and Wheater, 2002).

3.3.3 Modelling strategy

Following Yan et al. (2002) the modelling process started with the development of a baseline model composed of basic factors influencing rainfall variability such as geographical and seasonal factors. The progressive addition of the terms followed a perceived order of importance and insignificant factors were deleted throughout the procedure to keep the model manageable. At each stage, Pearson residual means (Section 3.3.2) were used to check the baseline structure of the model and suggest directions for improvements. The result of this initial stage was a model describing seasonal and regional variation in southern African rainfall variability.

Subsequently, time varying climate factors, which have been reported as influencing rainfall variability over the subcontinent as summarised in Section 1.2.1, were added. Initially, the statistical significance of each factor was examined, individually adding the

factor to the model. The candidates were then added to the model in order of significance, along with a predefined set of interactions (two degrees of polynomials to represent regionality and one to represent altitude, annual cycle to describe seasonality and autocorrelation; refer to the following sections for a description of such covariates). Finally, the terms were pruned in case of insignificance. The procedure was then repeated with the remaining factors, until no remaining factors were significant at the 1% level. Because blocks of terms rather then single ones were considered simultaneously, statistical significance was assessed using likelihood ratio tests, adjusted for inter-site correlation using the methodology described in Chandler and Bate (2007). That methodology also corrects for possible failure of assumption of constant CV (see Section 3.3.1).
The 1% threshold for determining significance is more stringent than is commonly applied in climatological studies. This has been done in an attempt to avoid overfitting, since the dataset is so large that even small effects that are practically negligible (and hence of little interest) may appear statistically significant at more conventional levels. In the procedure outlined above, the significance of each climate index was tested by adding the main effect along with seasonal and regional interactions, to allow flexibility in representing the structures of those effects.

3.3.4 Regional rainfall structure

To describe the climatology of the region, covariates representing regional and seasonal variation have been included in the models. Systematic regional variation is represented using Legendre polynomial transformations of latitude, longitude and altitude (the latter is the median value of the grid cell) (Abramowitz and Stegun, 1965, see Figure 3.2). This ensure that the covariates are, as far as possible, independent (Chandler, 2005).
Legendre polynomial are defined, for interval of x $\in [a, b]$, as:

$$P_1 = \frac{2x - (a + b)}{b - a};$$

where, if $x = a$ then $P_i = -1$ and if $x = b$ then $P_i = 1$.

$$P_2 = \frac{3P_1^2 - 1}{2};$$

$$P_3 = \frac{(5P_1P_2) - (2P_1)}{3};$$

$$P_4 = \frac{(7P_1P_3) - (3P_2)}{4}$$

Despite the use of four degrees of polynomials to represent regionality in the amount model, initial modelling attempts revealed that they were not sufficiently flexible to represent the effect of the Namib Desert, which is a relatively local but important feature. Therefore a 'Namib' variable, defined at each grid cell to take values 0, 0.3, 0.6 or 1 depending on the proximity to the Namib Desert, has been added to the model. Following the preliminary analysis, the grid cells that are characterised by the driest climate, and assumed to encompass the Namib, are coded as 1, 0.6 and 0.3 if in the nearest proximity or at the boundary of the desert. Finally those that are uninfluenced by the desert are coded as 0. Although this definition will not change under a future climate change, this is not assumed to be a major limitation of the model when used for application in the current century. Indeed, no precipitation trend and little precipitation variability has been found in records for the past hundred years.

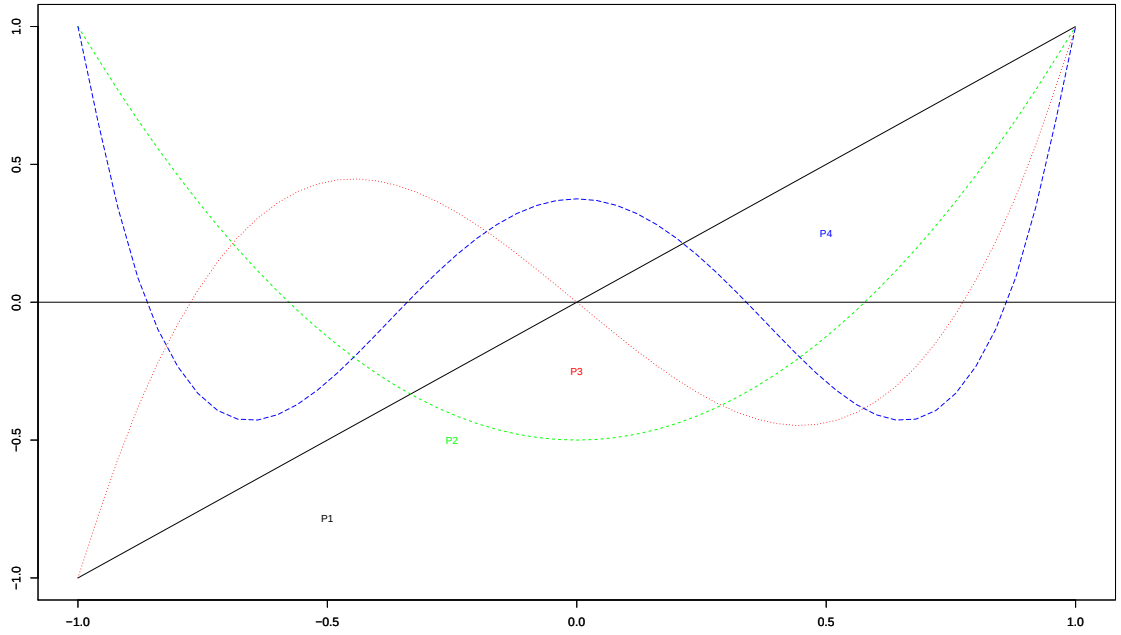
3.3.5 Seasonal rainfall structure

Periodic signals representing seasonality throughout the region are described using a Fourier series of sine and cosine covariates (Chandler, 2005) defined as:

$$\cos\left(2n\pi\frac{t}{S}\right);\sin\left(2n\pi\frac{t}{S}\right)$$

where S is the number of time units in a complete seasonal cycle (e.g. 12 if t is measured in months) and $n = 1$ represents the fundamental frequency of the annual cycle; increasing n gives additional harmonics (Chandler and Scott, 2011, see Figure 3.3). Significant terms representing the annual cycle plus two harmonics, to capture the length of the wet season, were added to the model.

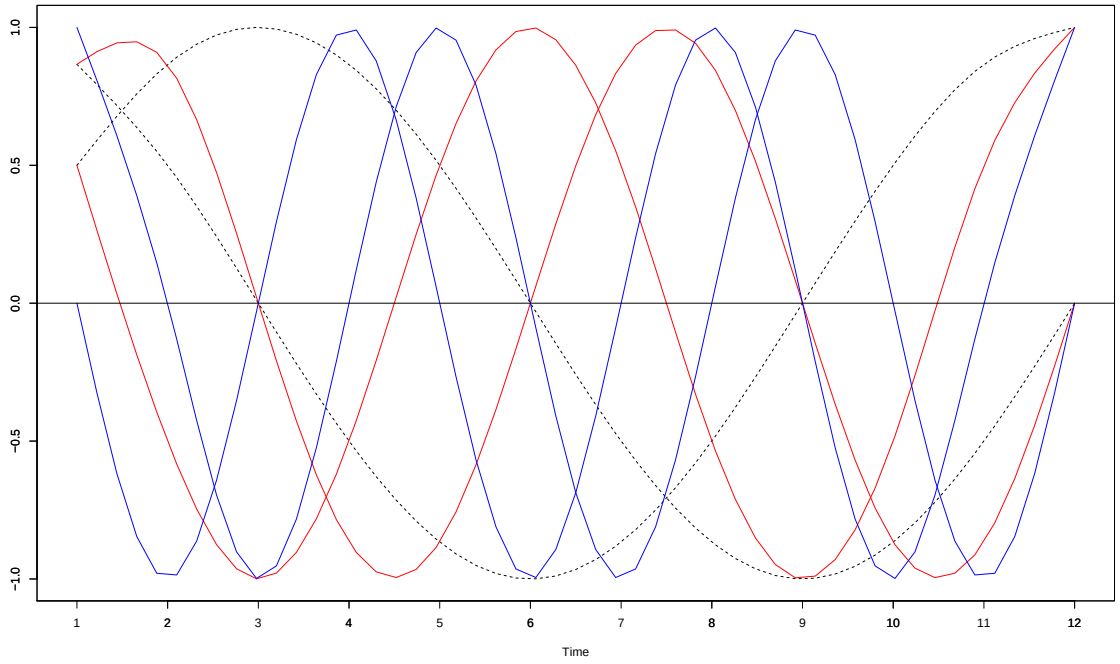
FIGURE 3.2: Illustration of the first four Legendre polynomials for interval of x $\in$ $[-1, 1]$.



3.3.6 Temporal and spatial dependency structure

Common issues, such as temporal and spatial dependencies in climatic data sets, are taken into account. In order for the models to capture the persistence of periods of below/above normal precipitation, terms representing lagged rainfall have been included. For the occurrence model the corresponding covariate is a binary (0/1) variable indicating whether or not rainfall occurred in the previous month at the same location. In the amount model, this covariate is replaced by log(1+previous month's rainfall). The spatial dependence is considered as though the data for each location are independent and later adjusting the standard error of the estimated parameters for the dependence between neighbouring sites (see methodology in Chandler and Bate, 2007).

FIGURE 3.3: Illustration of Fourier basis for interval $S = 12$ and t measured in months. Dashed, black line representing the annual cycle; red and blue lines representing two harmonics.



3.3.7 Simulations

As well as studying the effect of climate covariates upon precipitation, GLMs enable us to simulate rainfall time series conditional on the selected covariates. To simulate rainfall for a single month, the first step is to use the fitted models to calculate the probabilities of rainfall occurrence and the expected rainfall intensities at all locations for that month. Binary variables representing precipitation occurrence can then be generated according to the calculated probabilities; then, for each wet location, an amount of precipitation can be sampled from the appropriate gamma distribution. As explained in Section 3.3.6, in the occurrence and amount models the temporal dependence in the response variable is accounted via the previous month rainfall covariates. Additionally, in the case of spatially distributed observations, nearby locations tend to have high correlation, and that can be particularly true for climate variables where a common weather system may influence adjacent sites. Since in the models there are no variables

representing the spatial dependence, the same should be specifically accounted during the simulation procedure. One way to achieve this (Yan et al., 2006) is via a model for the inter-site correlations between Anscombe residuals (Section 3.3.2).

Typically, the correlation between two sites at locations (Lat_1, $Long_1$) and (Lat_2, $Long_2$), will depend on the distance between them. The distance (d) among all pair of sites at longitudinal and latitudinal locations was calculated in nautical miles from Roy and Clarke (1988):

$$\mathrm{d} = \left(\frac{60 \times 180}{\pi}\right) \cos^{-1}\left[\sin\left(Lat_1 \times \frac{\pi}{180}\right) \sin\left(Lat_2 \times \frac{\pi}{180}\right)\right] + \left[\cos\left(Lat_1 \times \frac{\pi}{180}\right) \cos\left(Lat_2 \times \frac{\pi}{180}\right) \cos\left((Long_2 - Long_1) \times \frac{\pi}{180}\right)\right]$$

The distance between pairs of sites was computed with the above formula and used to fit a spatial correlation model for the Anscombe residuals for the amounts model.

As described in Yang et al. (2005), a spatial field of correlated rainfall intensities can be generated from the Anscombe residual correlations along with the shape parameter of the gamma distribution and the expected rainfalls from the amounts model. Spatial correlation is not considered in the occurrence model due to the complexity in the generation of binary sequences versus the benefit expected to be modest. (Yang et al., 2005). Indeed, as a result of the few dry months being mostly concentrated in the same region and season, the spatial correlation is already largely accounted for by the covariates representing regionality and seasonality.

3.4 Modelling results

3.4.1 Modelling the spatial and seasonal variation

3.4.1.1 Occurrence model

In the baseline occurrence model (see full model specification in Appendix A), Legendre polynomials of degree two for altitude and degree three for latitude and longitude were used to describe respectively the variation of precipitation occurrence with altitude

and spatial location. Seasonality and temporal autocorrelation were represented using a Fourier basis and binary covariates respectively, as described in Sections 3.3.5 and 3.3.6. Besides the individual factors, the baseline model also includes 2- and 3-way interactions between regional, seasonal and autocorrelation terms. The interpretation of the 3-way interactions is that the seasonal variation in autocorrelation itself varies over the subcontinent.

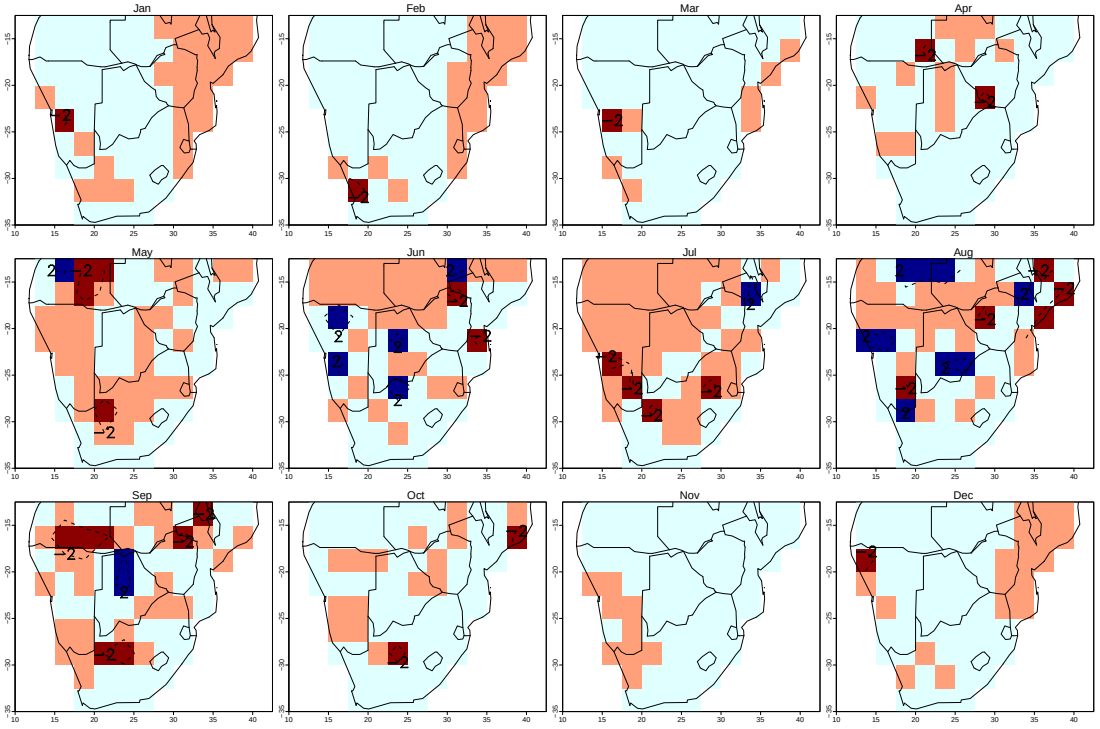
Standardised mean Pearson residuals (Figure 3.4) were computed for each month of the year at each grid cell. Grid cells in red correspond to values < -2 and in blue to values > 2, where systematic biases are found. The cells in light cyan and salmon coincide to values within the range -2/0 (salmon) and 0/+2 (cyan), not significantly different from zero under the assumption that the model is correct (see Section 3.3.2). The plotted residual means indicate that the wet season (October-April) is extremely well represented throughout the sub-continent. During the remaining four months, in the dry season, from May to September, large standardised mean residuals occur mainly in the central part of the subcontinent, over Angola, Namibia, Botswana, Mozambique and central South Africa. These are the areas where the model, which aims primarily to describe large-scale structures, is unable to capture relatively localised features characterising the transition between highland and plain, desert and savannah. However, this is not considered too problematic considering that the overall aim of the current research is the study of periods of water stress and characteristics during the rainy season, since it is here that a shortage of water may lead to potentially disastrous impacts.

3.4.1.2 Amounts model

The basic covariates in the best fitting amounts model are similar to those for the occurrence model. Some additions have been made such as degree four Legendre polynomials for latitude and longitude, a degree three Legendre polynomial for altitude and a term representing the Namib Desert local drying effect. In addition, autocorrelation is represented via a different transformation of the previous month's rainfall as described in Section 3.3.6.

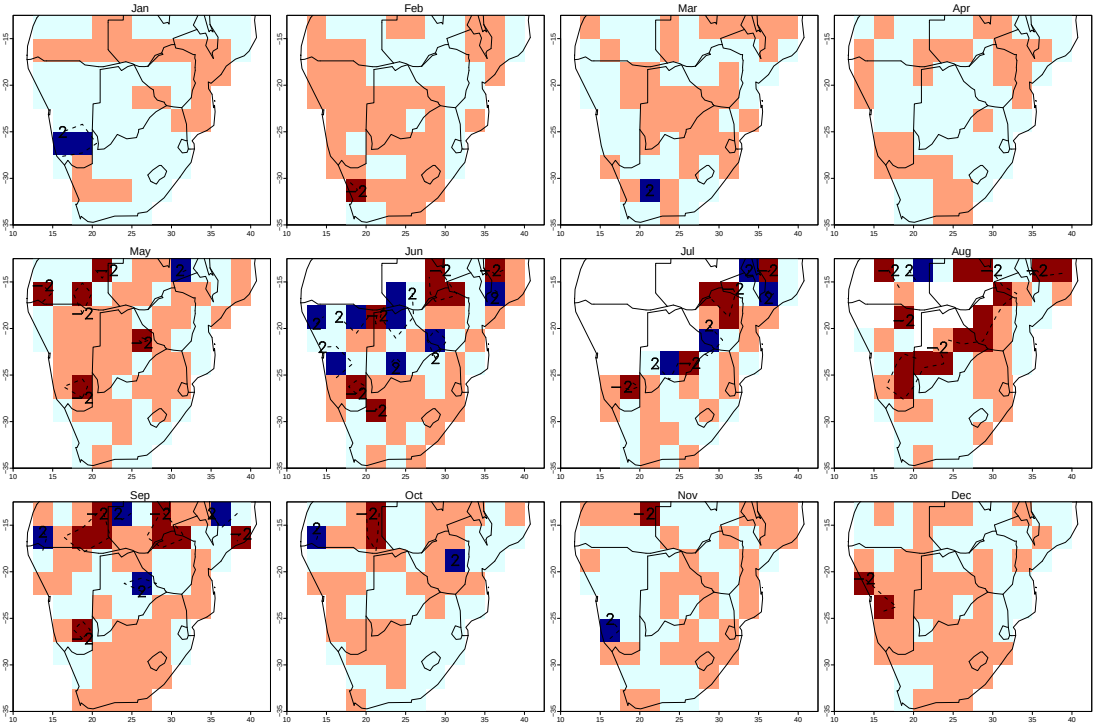
Once again the Pearson residual means (Figure 3.5; same colour scale as in Figure 3.5) indicate that the regional structure is well captured during the wet season (from October

FIGURE 3.4: Pearson residual mean for the occurrence model. Values between ±2 are represented in salmon (−2/0) and cyan (+2/0); values < -2 (corresponding to overestimation by the model) in darkred; and values > 2 in darkblue.



to April). On the other hand, in the remaining five months several grid cells have mean residuals that are significantly different from zero. Upon further investigation however, these grid cells correspond to regions with very few wet months over the period used to fit the model. In particular the highest residual is found in northern Botswana in June where only two out of the 50 months experienced rainfall. In this case, since the sample size is so small the normal approximation, which underlies the classification of a standardised mean residual outside the range (-2,2) as 'problematic', is likely to be extremely inaccurate. Moreover, in all cases with large standardised mean residuals, the maximum recorded rainfall was below 20 mm and the predicted amount below 2 mm. Once again therefore, the poor performance is relatively unimportant in practical terms. Blank cells in the plots correspond to grid cells for which no rainfall was experienced during the month in question throughout the period used for model fitting.

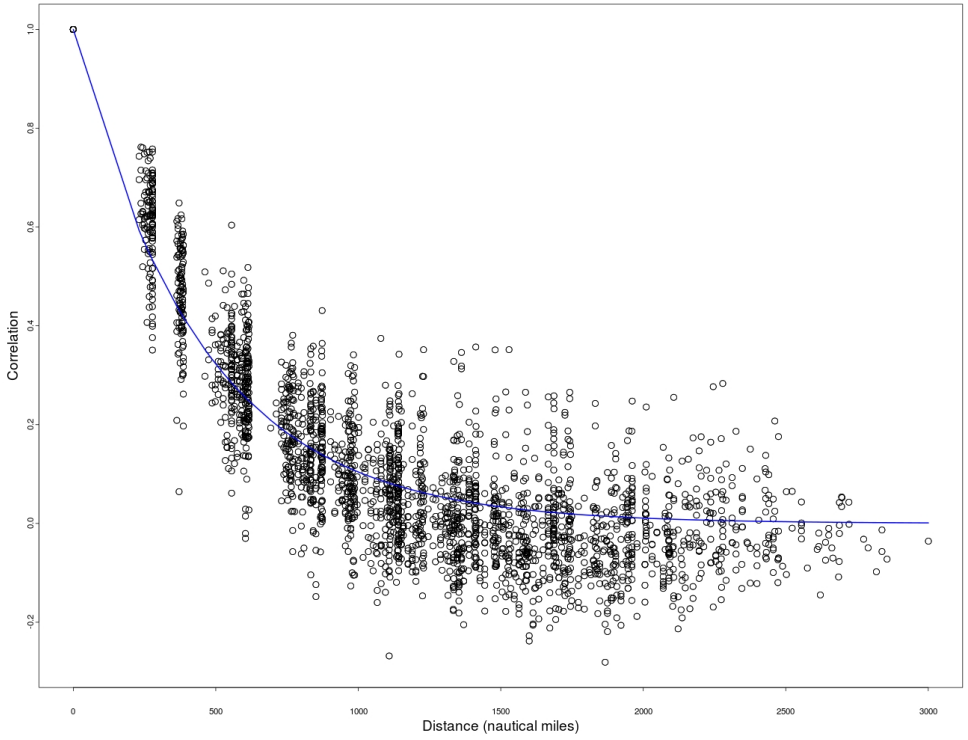
FIGURE 3.5: Pearson residual mean for the amounts model. Values between ±2 are represented in salmon (−2/0) and cyan (+2/0); values < -2 (corresponding to overestimation by the model) in darkred; and values > 2 in darkblue.



3.4.1.3 Spatial correlation model

During the simulation process (see Section 3.3.7), to account for the spatial dependence in rainfall amounts, correlations were estimated among pairwise complete observations; spatial correlation models were then fitted to these correlations using non-linear least squares. An exponential correlation model, showing correlation decay with increased distances, was found to capture the overall shape (Figure 3.6).

FIGURE 3.6: Exponential (in blue) correlation function fit to the pairwise correlations in the Anscombe residuals according to their inter-site distances.



3.4.2 Effects of climate indices

Having developed a baseline model, the potential effects of various climate indices were explored. Candidate indices were initially identified from the literature, as reported in Section 1.2.1 and Table 3.1 in the current chapter.

From the list in Table 3.1, six final indices were selected progressively according to their perceived order of importance and significance. Interactions were also considered and finally the insignificant terms were removed throughout the modelling exercise (see

Section 3.3.3). The first four indices were added in both the occurrence and amounts model, whereas the final two were selected only for the amounts model. The retained indices, in the order of significance, were as follows:

a) The first principal component of the relative humidity field over southern Africa. The loading pattern of the index is characterised by a circular feature over the central area of southern Africa (Figure 3.7). High values of the index correspond to periods with reduced relative humidity in the central area as compared with increases in the surrounding areas. Given the close link between precipitation and atmospheric moisture, it is not surprising that this index was the most statistically significant of those considered;

b) El Niño represented by the El Niño index and considered in the literature as one of the most important teleconnections affecting southern African rainfall;

c) Southern hemisphere temperature index (SHT), as representative of warming trend;

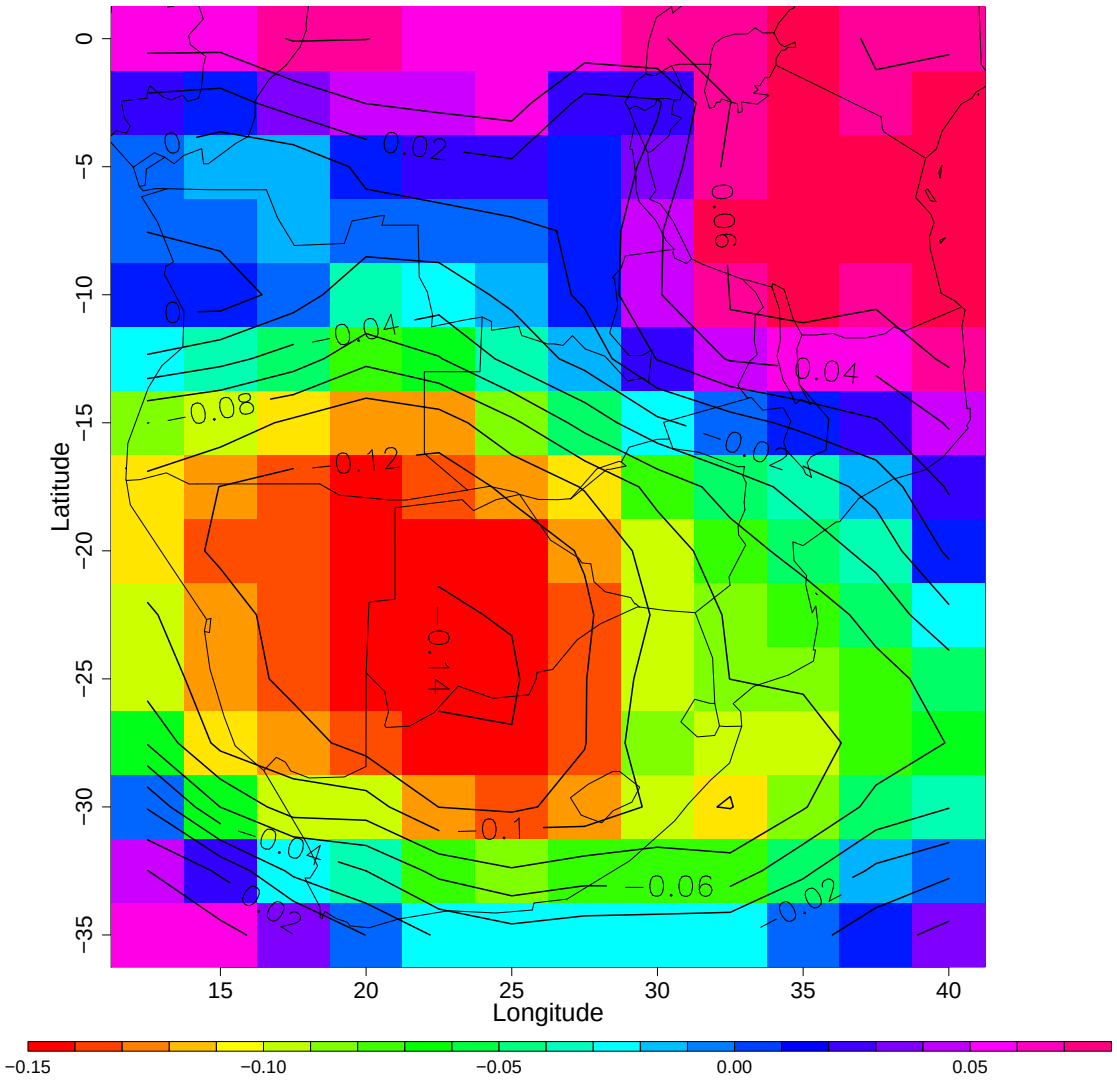
d) Two variables representing the spatially and temporally varying effect of the subtropical jet stream. The first is the standardised jet stream intensity, INT_t, see Section 3.2. The second is defined, for a location at latitude ℓ at time t, as

$$\text{POSEFF}_{\ell,t} = \max\left(0, \text{POS}_t - \ell\right) \ ; \tag{3.5}$$

where POS_t is the position of the jet stream (again, see 3.2.2 for definition). The variable POSEFF_t takes the value zero for locations to the north of the jet stream position at time t, and increases linearly with distance for locations to the south. As such, it can be regarded as a measure of the extent to which latitude ℓ is affected by mid-latitude weather systems at time t. The motivation for this definition is deferred until discussion of the amounts model below;

e) North-south gradient in South West Indian Ocean sea surface temperatures. This is defined as the standardised difference between average standardised SST anomalies in the north–west (-25°– -5°N, 60°– 85°E) and south–west (-40°– -25°N, 40°– 70°E) Indian Ocean;

FIGURE 3.7: First loading pattern for the PCA of standardised relative humidity field anomalies over southern Africa.



f) The Southern Annular Mode (SAM). The index used here is a mean sea level pressure (MSLP) difference between data of twelve stations in the Southern Hemisphere at both 40°S and 65°S (Table 3.1).

In order to avoid the results depending on the magnitude of the single indices, they all have been standardised separately for each month of the year, subtracting the monthly means and dividing by the standard deviations with respect to the period 1961-1990.

3.4.2.1 Occurrence Model and effects of climate covariates

Having included the first four indices listed in the previous section into the occurrence model, none of the remaining indices in Table 3.1 was significant at the 1% level. The final model included 139 terms of which 23 are main effects, 91 are 2-way interactions and 24 are 3-way interactions. Such a large number of terms may seem, at first sight, excessive, but the model was developed to represent seasonal and temporal rainfall variability over a large and topographically complex region. Indeed, 103 of the 139 terms relate to the non-parametric representation of regional variation. As Yan et al. (2006) pointed out, a model of this size for the representation of almost 50,000 data points is parsimonious if compared to a conventional approach of working separately with monthly anomalies at each of the 72 grid cells, since the computation of such anomalies itself involves the estimation of $72 \times 12 = 864$ means at the outset.

The use of the Legendre polynomial in representing the regional structure of precipitation may be considered prescriptive, although this particular methodology is designed to provide a flexible representation of structure that are expected to be smooth. Alternative data driven representation, such as those based on non-parametric regression techniques (Bowman et al., 2009), may give more accurate results at the expenses of greater computation intensity.

To visualize the modelled effects of the various climate indices, their contributions to the linear predictor in equation 3.1 were considered. Specifically, for each grid cell and each month the value of this linear predictor was calculated with all remaining time-varying covariates set to their average values. The value of each index in turn was later increased by one unit and the linear predictor recalculated. The change in the linear predictor represents the effect of a 1-unit increase of the index upon the log odds of rainfall. Because of interactions involving seasonal and regional covariates, this effect itself varies seasonally and regionally.

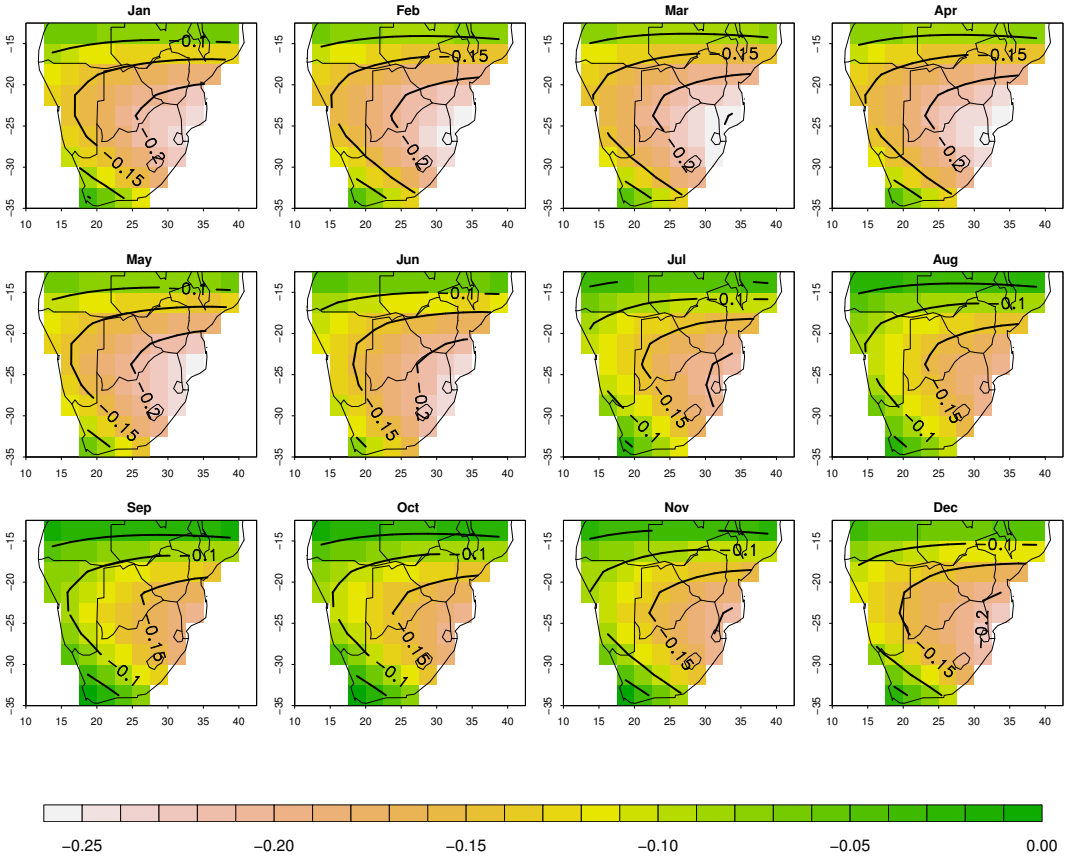
The modelled effects are shown in Figures 3.8 to 3.10, and can be summarised as follows:

a) Relative Humidity Index

   Figure 3.8 shows the effect on precipitation occurrence of a 1-unit increase in the relative humidity index defined by the first loading pattern over the subcontinent. This show that an increase in the derived index is associated with a reduction in

rainfall occurrence over the eastern and central area and a corresponding increase elsewhere, particularly during the austral winter. Because of the way the index has been defined, increased index values corresponds to reduced humidity in the central area. The pattern found here is consistent with the first principal component loading pattern described above (Figure 3.7), although shifted slightly eastwards.

FIGURE 3.8: Modelled effect of Relative Humidity upon rainfall occurrence. Contours represent the effect of a 1-unit increase in the Relative Humidity Index on the log odds of monthly rainfall occurrence.

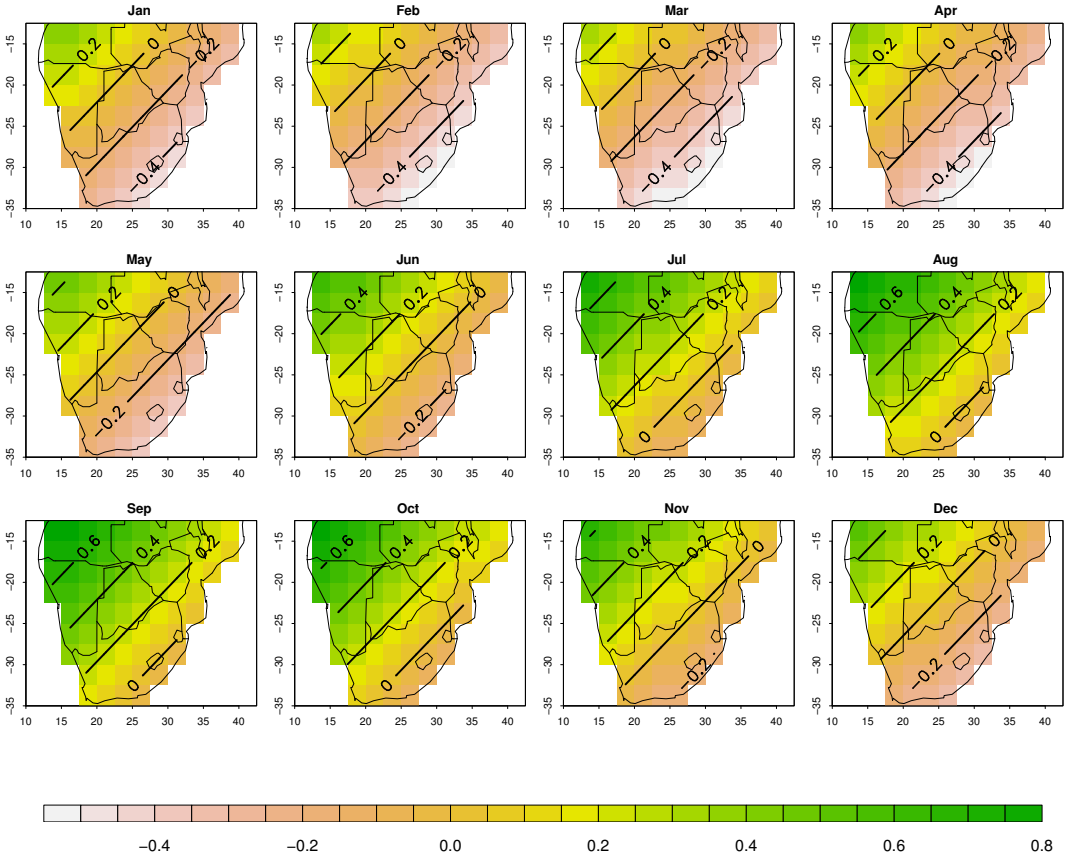


b) El Niño Index

Figure 3.9 shows that according to the fitted occurrence model, an increase in the El Niño index is associated with a reduction in the probability of rainfall occurrence during the austral summer over much of the subcontinent, especially over the south-eastern part of it. A clear south-east north-west gradient is also visible, with increases in occurrence in the northwest particularly from July to December. By negatively

affecting the probability of rainfall over the southeast of the subcontinent, El Niño has a shortening effect upon the length of the rainy season. The effect of El Niño can therefore be summarised as reducing the amplitude of the seasonal cycle in rainfall occurrence in the southeast of southern Africa, mostly associated with a reduction in the probability of rainfall during the wet season in accordance with previous studies (Section 1.2.1).

FIGURE 3.9: Modelled effect of El Niño upon rainfall occurrence. Contours represent the effect of a 1-unit increase in the El Niño Index on the log odds of monthly rainfall occurrence.

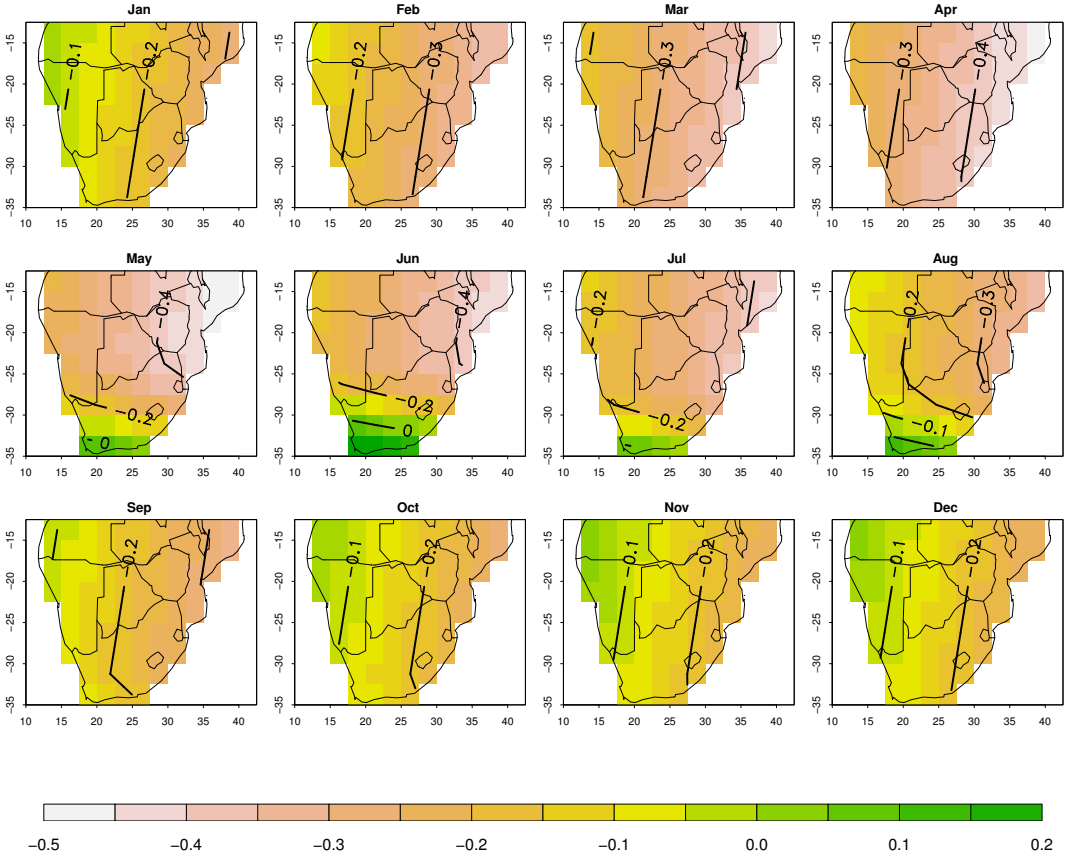


c) Southern Hemisphere Temperature Index

Since the ocean tends to filter out high-frequency variation, SHT is a good index of underlying changes in global temperature. Its effect can therefore be interpreted as a global warming signal (Ster and Kaufmann, 2000). Figure 3.10 shows the modelled effect of a 1-unit increase in SHT which, in broad terms, corresponds to a reduction

in the probability of precipitation occurrence throughout the year, except for the winter southern Cape. This is consistent with the conclusions of Bates et al. (2008): higher SHTs are very likely to lead to an increase in the proportion of land surface in drought at any one time. In addition, a tendency for future drying in continental interiors during summer, especially in the subtropics, low and mid-latitudes has been projected and, according to Shongwe et al. (2009), with increasing temperature the summer precipitation may be reduced, mainly visible in a later onset, and result in an eastward extension of desert areas in southern Africa. Again, Figure 3.10 supports this since the modelled decreases in precipitation occurrence are greater in the east.

FIGURE 3.10: Modelled effect of Southern Hemisphere mean Temperature upon rainfall occurrence. Contours represent the effect of a 1-unit increase in the Southern Hemisphere mean Temperature Index on the log odds of monthly rainfall occurrence.



d) Jet Stream effects

Due to its complexity, the modelled effect of a strengthening and displacement of

the subtropical jet stream cannot be easily shown with maps similar to the previous ones. Broadly speaking, inspection of the model coefficients suggests that a stronger jet stream intensity is associated with an increase in rainfall occurrence, particularly in locations south of the current jet stream position.

3.4.2.2 Amounts Model and effects of climate covariates

Besides the basic structure, the final amounts model also contains terms representing the effect of El Niño, SHT, relative humidity and jet stream intensity and position as for the occurrence final model. In addition, two further climate covariates were found to be significant: the South West Indian Ocean (SWIO) SSTs gradient and the Southern Annular Mode (SAM).

More interactions were found to be significant than for the occurrence model, bringing the total number of terms to 244 (33 main effects, 186 two-way and 24 three-way interactions) of which 189 are related to the representation of regional variation and are needed to represent the complex topographic controls on rainfall in the area.

As for the occurrence model, the effect of each climate index can be visualised by considering the change in the linear predictor resulting from a unit increase in the index. However, since a log link function has been used (see equation 3.2), a change of (say) δ in the linear predictor corresponds to multiplying the expected monthly rainfall by exp (δ). We present the estimated multiplicative effects here, for ease of interpretation.

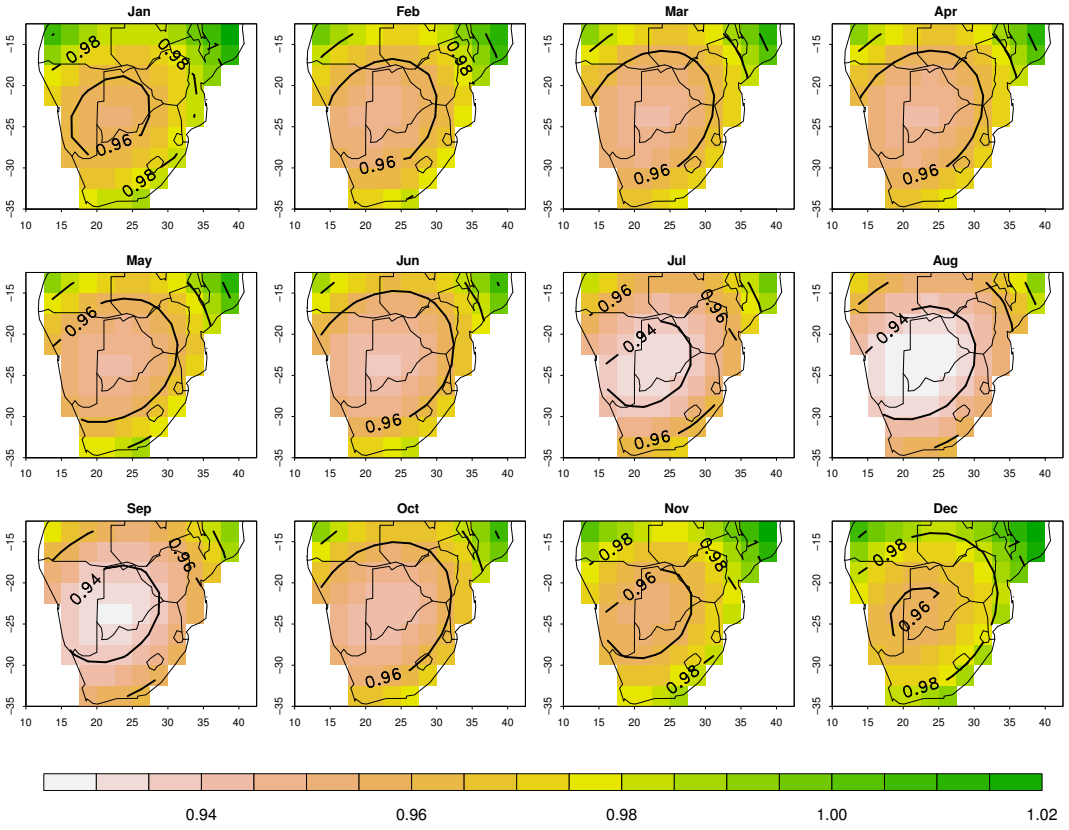
a) Relative Humidity

   The pattern shown in Figure 3.11 appears highly consistent with the loading pattern used as index. The main feature is centrally situated over central southern Africa, in contrast with the effect modelled by the occurrence model, where the main effect was shifted towards the Indian Ocean (Figure 3.8). A 1-unit increase in the index leads to a reduction in the expected monthly rainfall over the central area of the subcontinent, and to an increase around the edges.

b) El Niño

   As with the rainfall occurrence model, El Niño shows clear seasonally and locally varying effects (Figure 3.12). The picture shows the effect of 1-unit increase of El

FIGURE 3.11: Modelled effect of Relative Humidity upon rainfall amount. Contours represent the multiplicative effect of a 1-unit increase in the Relative Humidity Index on the expected monthly rainfall amount.



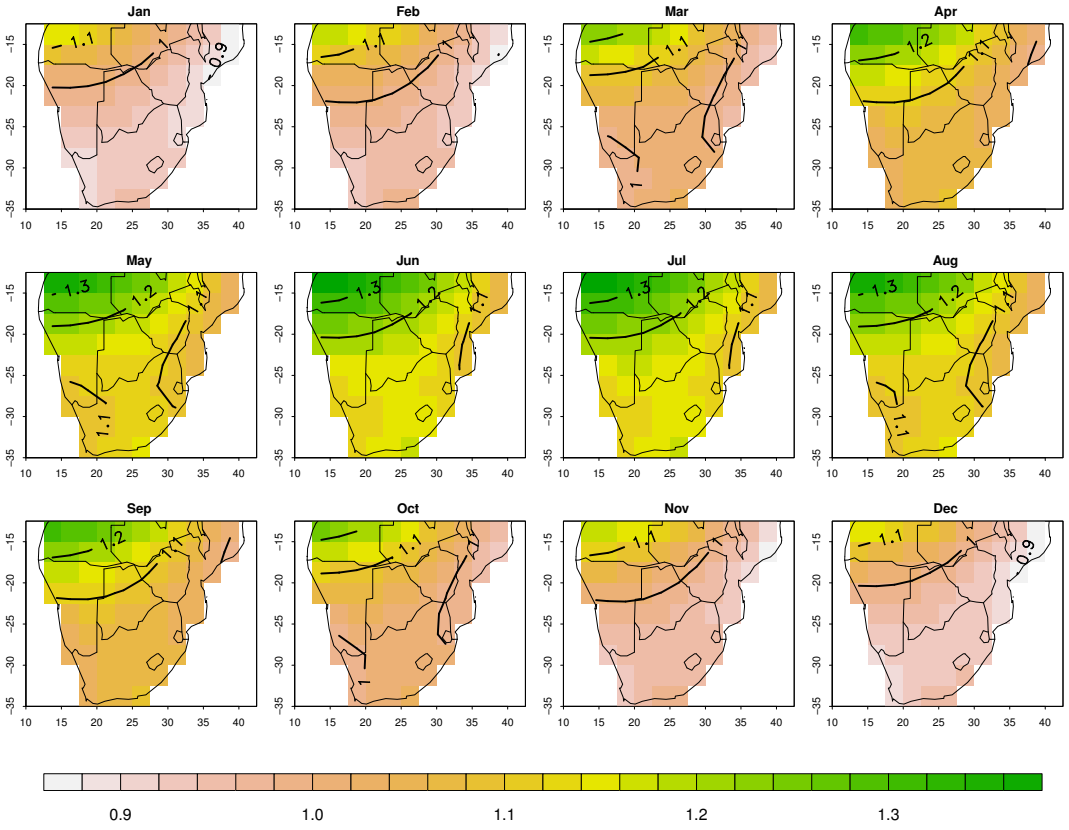
Niño index conditional on the average values of the relative humidity with which it interacts. The strongest impact is seen in the central and southern area of southern Africa and in the austral summer: here, an increase in the El Niño index is associated with a reduction in mean rainfall, in agreement with previous studies (e.g. Nicholson and Kim, 1997).

c) Southern Hemisphere Mean Temperature

Figure 3.13 shows the modelled effect of a unit increase in the SHT index conditional on the average values of the relative humidity index (since an interaction between the relative humidity and SHT indices was found to be significant).

As with the rainfall occurrence model, the modelled effect of an increase in SHT corresponds to an overall reduction in expected precipitation throughout the year,

FIGURE 3.12: Modelled effect of El Niño upon rainfall amount. Contours represent the multiplicative effect of a 1-unit increase in the El Niño Index on the expected monthly rainfall amount.



especially over Namibia and Botswana (Figure 3.13) which are already characterised by dry climates.

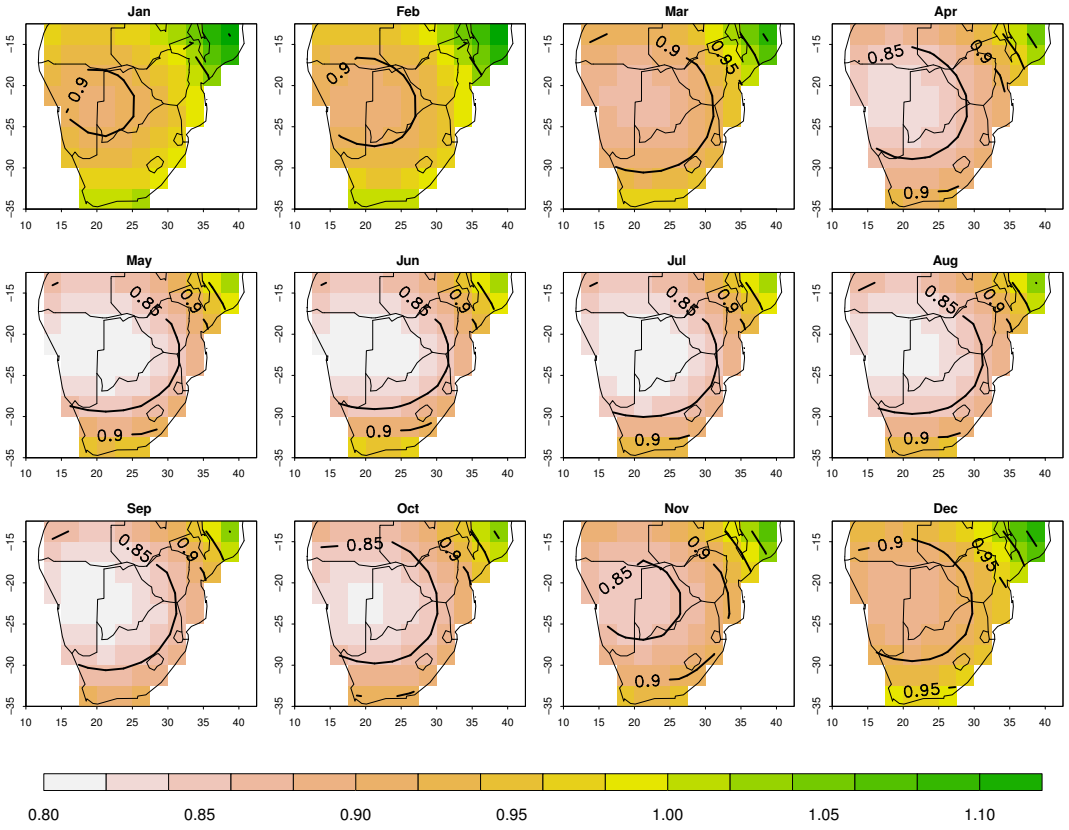
d) Southern Annular Mode

Figure 3.14 shows the modelled effect of a unit increase in the SAM index. An increase in the SAM index corresponds to enhanced westerlies in the southern oceans with a large scale transfer of momentum and energy in the form of heat and moisture from the tropical to middle latitudes. The effect agrees with the results of Gillett et al. (2006): the positive phase of the SAM is associated with an increase in precipitation intensity over most of South Africa.

e) South West Indian Ocean SSTs

The effect of the South West Indian Ocean dipole index is shown in Figure 3.15.

FIGURE 3.13: Modelled effect of Southern Hemisphere Temperature upon rainfall amount. Contours represent the multiplicative effect of a 1-unit increase in the Southern Hemisphere Temperature Index on the expected monthly rainfall amount.



For the summer, this shows a similar pattern to the one found in Washington and Preston (2006): an increase in the north-south SST gradient is associated with an increase in the expected monthly rainfall during late summer months (mainly from February to May). However, the opposite is true for the remaining months.

f) Jet stream intensity and position

The effect of the jet stream is represented, as for the occurrence model, via the variables $\text{POSEFF}_{\ell,t}$ and INT_t defined earlier. The variable $\text{POSEFF}_{\ell,t}$ was defined by considering that the effect of the jet stream intensity INT_t would probably vary spatially depending on the position POS_t of the jet stream at time t (Section 3.2) relative to the location of interest. To gain some insight into the nature of this spatial variation, residuals were computed from the amounts model containing all of the factors listed above. For each grid cell, the relative position of the jet stream (in

degrees north of the grid cell centre) was then computed at each time point. The residuals were then grouped according to the values of this relative position and, for each group in turn, were regressed upon the jet stream intensity. Figure 3.16 shows the resulting regression coefficients for each group (solid line). This shows clearly that the regression coefficients are almost zero for groups corresponding to negative relative jet stream positions (i.e. for groups where the jet stream is to the south of the location of interest), and increase roughly linearly for positive values of the relative position. The high values of the regression coefficients for values south of 40°S is due to availability of only few observations and the high sampling variability. For such reason, and because those locations are located over the Southern Ocean, such high coefficients have not been taken into consideration. Thus the regression coefficient of jet stream intensity INT_t can itself be represented roughly as a linear function of the variable $\text{POSEFF}_{\ell,t}$, defined at equation 3.5. This structure can be incorporated directly in a GLM: it is represented by the interaction between INT_t and $\text{POSEFF}_{\ell,t}$. The definition of $\text{POSEFF}_{\ell,t}$ is intended to mimic this pattern. The effect found in the current study in consistent with that expressed by Richard et al. (2001) (Section 1.2.1) where a northward displacement of the jet stream could be linked to a northward displacement of the cyclonic system tracks.

FIGURE 3.14: Modelled effect of Southern Annular Mode upon rainfall amount. Contours represent the multiplicative effect of a 1-unit increase in the Southern Annular Mode Index on the expected monthly rainfall amount.

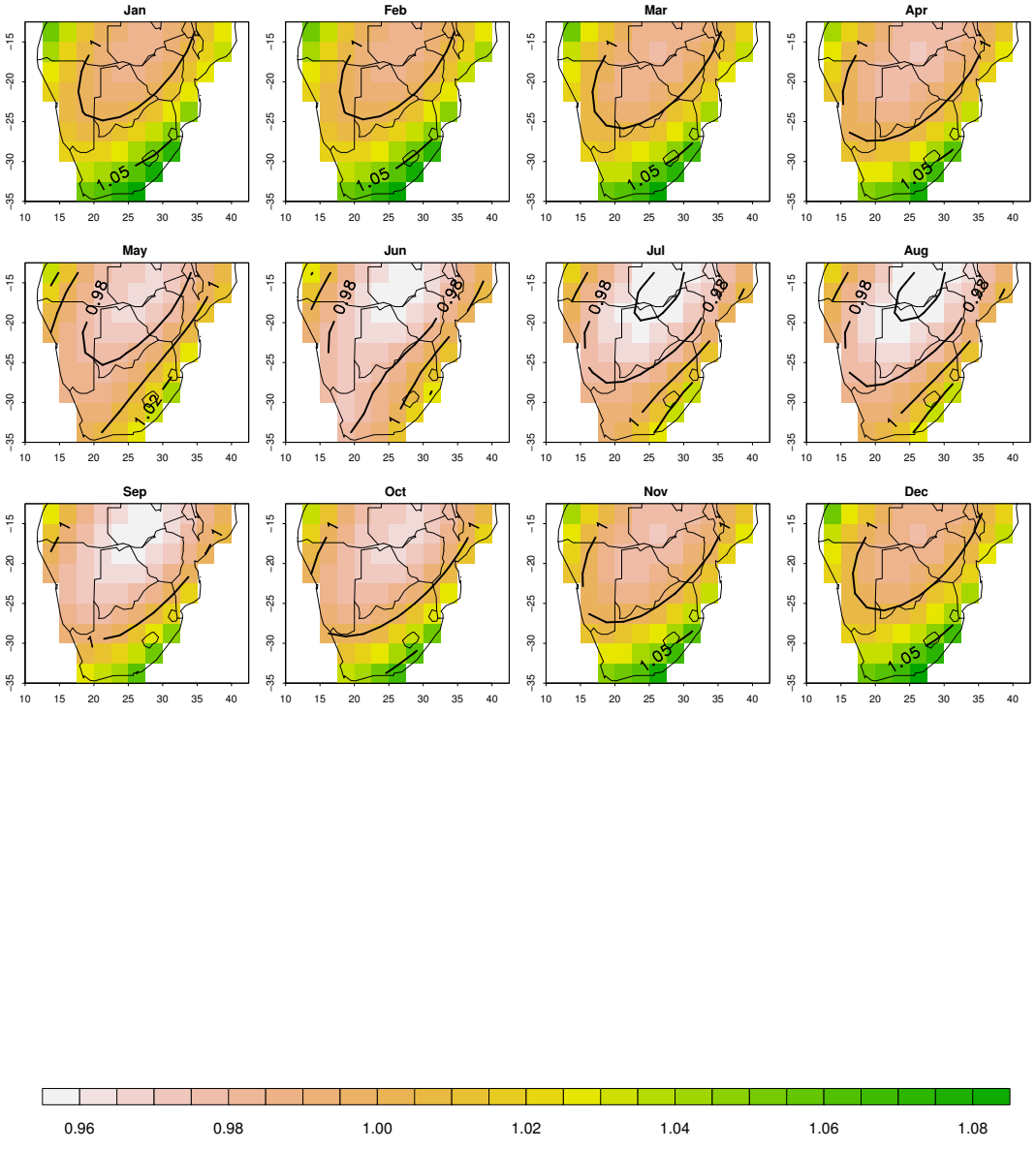


FIGURE 3.15: Modelled effect of South West Indian Ocean dipole index upon rainfall amount. Contours represent the multiplicative effect of a 1-unit increase in the South West Indian Ocean Index on the expected monthly rainfall amount.

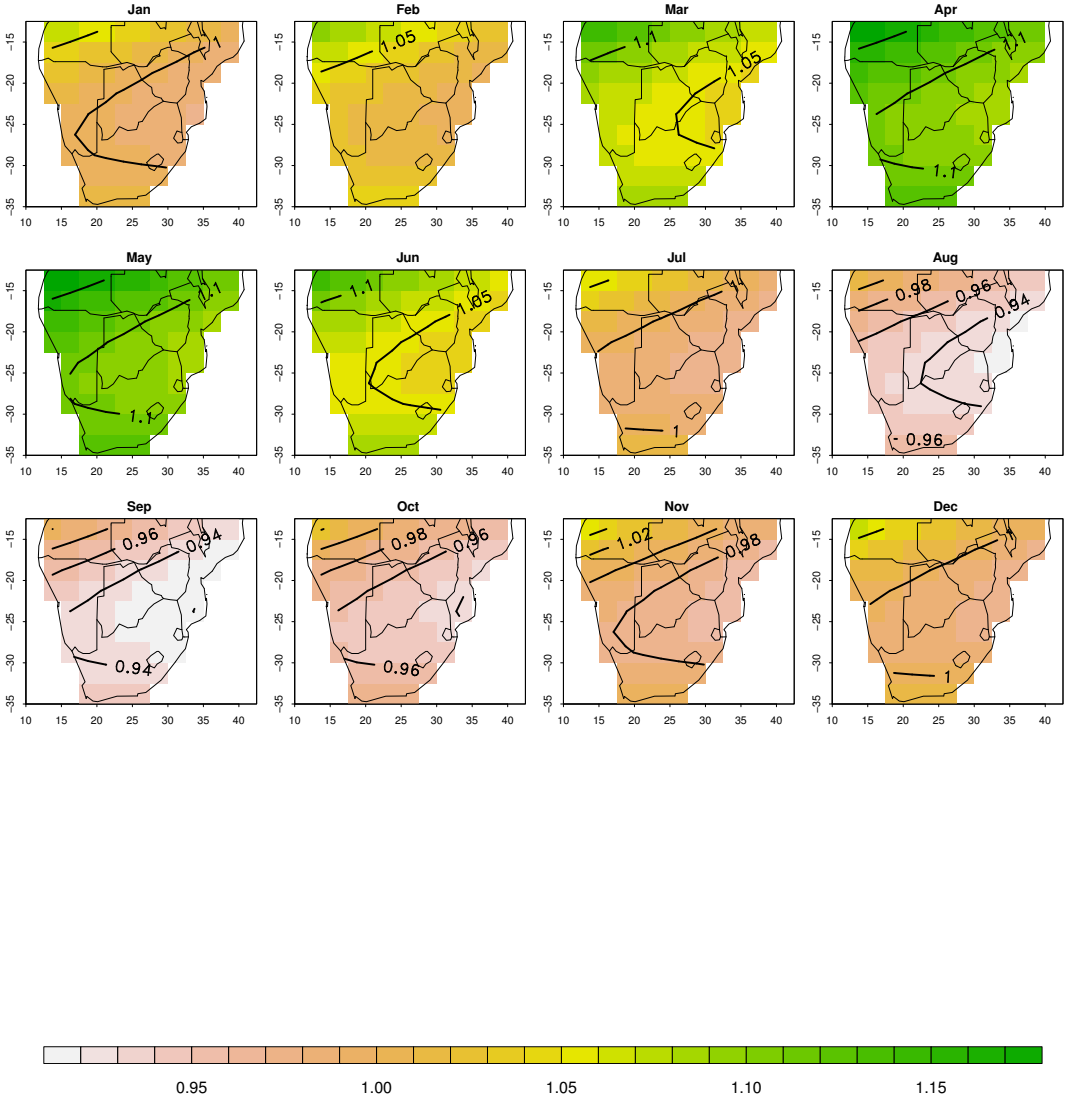
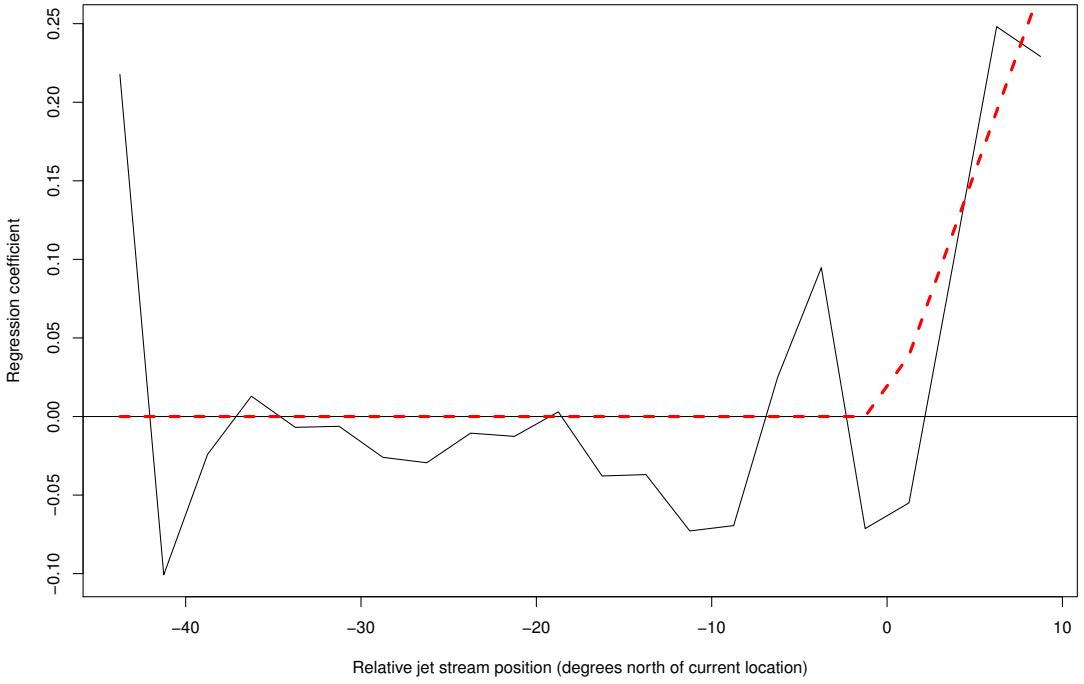


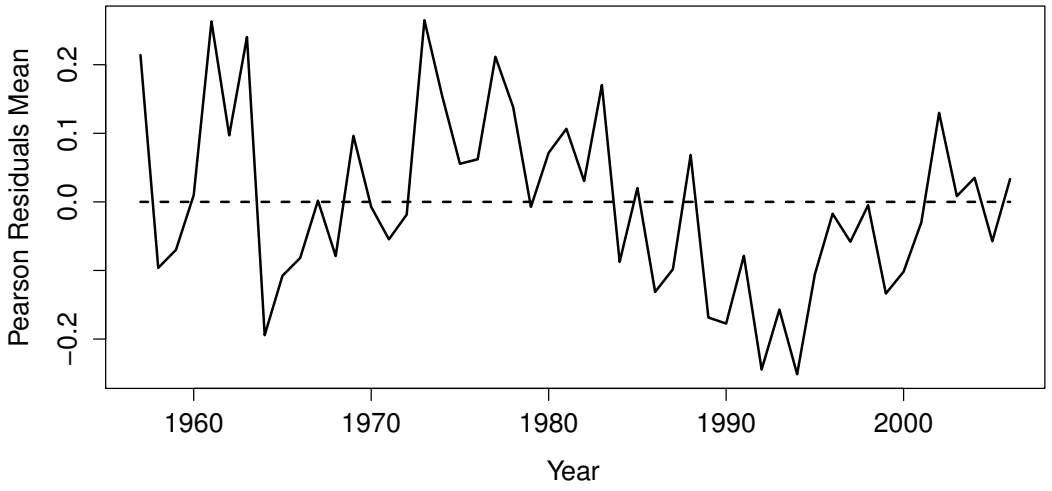
FIGURE 3.16: Definition of jet stream position variable $\text{POSEFF}_{\ell,t}$: coefficients from regression of residuals of penultimate rainfall amounts model upon jet stream intensity INT_t, plotted separately for different relative latitudinal jet stream positions (black solid line) and fitted values (red dashed line).



3.4.3 Final checking

As an additional check for unexplained structure in the final amounts model, annual mean Pearson residuals for the whole region were computed and are shown in Figure 3.17. Although no systematic trend is visible from the plot, there are runs of predominantly positive (e.g. during the 1980s) and negative (during the 1990s) values. This suggests the presence of interdecadal variability in precipitation over southern Africa which is not associated with any of the covariates in the fitted model, or with any of the other covariates considered (e.g. AMO and PDO, which are often considered as dominant modes of decadal variability; McCabe and Palecki, 2006).

FIGURE 3.17: Annual time series of mean Pearson residuals for the final rainfall amounts model.



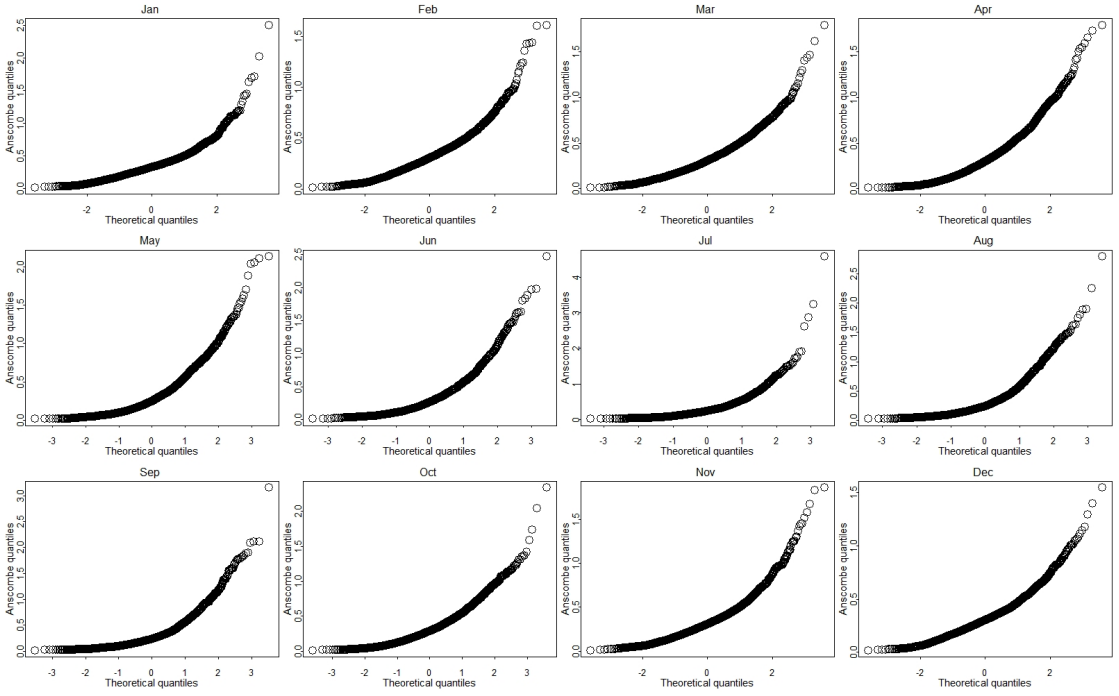
To check if the probability structure of the model is correct the Anscombe residuals are plotted (Section 3.3.2). If the assumption of a gamma distribution is correct the residuals should all come from an approximately normal distribution. In Figure 3.18 the Anscombe residual quantiles are plotted against equally spaced quantiles from a standard normal distribution. The nonlinearity in the points in the plots indicates a departure from normality. The problems in the lower tail are probably due to a breakdown in the normal approximation since the gamma distribution cannot yield negative

values while the normal can. The lack of fit in the upper tail can still be seen in the rainfall simulation (Section 3.5).

Although the Q-Q normal plot of the Anscombe residuals does not show a good fit of the chosen probability distribution, the final test of the model performance is the distribution of the simulated rainfall values (Section 3.5).

However, as stated by Chandler and Scott (2011), in a GLM, the parameter estimates depend on the underlying distribution only through the mean and variance, suggesting that the precise form of the distribution is relatively unimportant. Hence, in any situation where the variance appears related to the mean, models could be fitted as if the data were generated from the corresponding exponential family distribution.

FIGURE 3.18: Normal quantiles-quantiles plot of Anscombe residuals from the final amount model.



Notice that, according to the fitted models, several of the covariates identified in the literature (Section 3.2 and Table 3.1) do not have a significant effect on precipitation in the region. Thus, although there may be physical arguments for such links, the analysis here suggests that their effects are not relevant at a subcontinental scale, at least after other,

more important, factors have been accounted for. However, some of those factors appeared significant initially: this illustrates the importance of accounting simultaneously for all relevant factors in analyses of this type. A possible concern is that the effects of some potentially important climate covariates were masked by collinearities with other covariates in the model. To investigate this, variance inflation factors (Fox, 2002, pg. 216) were computed for each of the covariates considered. These provide measures of the increase in the variance of an estimated regression coefficient due to collinearities. Values in excess of 5 or 10 are generally considered to indicate problematic collinearity (Montgomery and Peck, 1992; Quinn and Keough, 2002): here, all of the values were less than 2 so we conclude that collinearity is not a major problem for the modelling exercise.

3.5 Model Simulations

In the modelling described above, all of the data were used to identify the overall model structure and identify the relevant predictors. Subsequently, a validation exercise was carried out in which the model was refitted to subsets of the data omitting successive decades in turn (1970s, 1980s and 1990s). Performance was assessed by simulating data for the periods not used in the fitting and comparing the results with the observations for these periods. This enables us to see whether or not a model using the selected predictors and calibrated using data from one time period, can reproduce the properties of precipitation over the subcontinent for a period not used for the calibration. The validation process is necessary to gain confidence in the model results later to be used for the simulation of future precipitation sequences.

The simulation was carried out as explained in Section 3.3.7, with the use of a spatial correlation model as described in Section 3.4.1.3.

The results are plotted as hydrological years, from July to June, in order to capture the wet season in full. In hydrological years the annual cycle is associated with the natural progression of the hydrological seasons and the beginning of the year represents the beginning of the rainy season. For the 50-year (49 hydrological years) calibration period and the three 10-year (9 hydrological years) validation subsets, 100 precipitation time series were simulated and the annual precipitation totals averaged over the entire

region. Thus, for each year of simulation a distribution of 100 simulated annual totals was obtained.

The plots in Figures 3.19, 3.20, 3.21 and 3.22 show the quantile bands corresponding to the 0-5th, 5th-10th, 10th-25th, 25th-50th, 50th-75th, 75th-90th, 90th-95th and 95th-100th percentiles obtained from the simulations, both for the entire period and for the three out-of-sample validation periods. In each plot, the observed time series is shown as a black line. If the simulations are realistic, the observed values should look as though they were sampled from the simulated distributions (Yang et al., 2005).

FIGURE 3.19: Observed time series (in black line) and rainbow range of annual regional average simulated distributions for 49 hydrological years (1958-2006) selected in the calibration process. Bands indicate the 0-5th, 5th-10th, 10th-25th, 25th-50th, 50th-75th, 75th-90th, 90th-95th and 95th-100th percentiles of the simulated distributions.

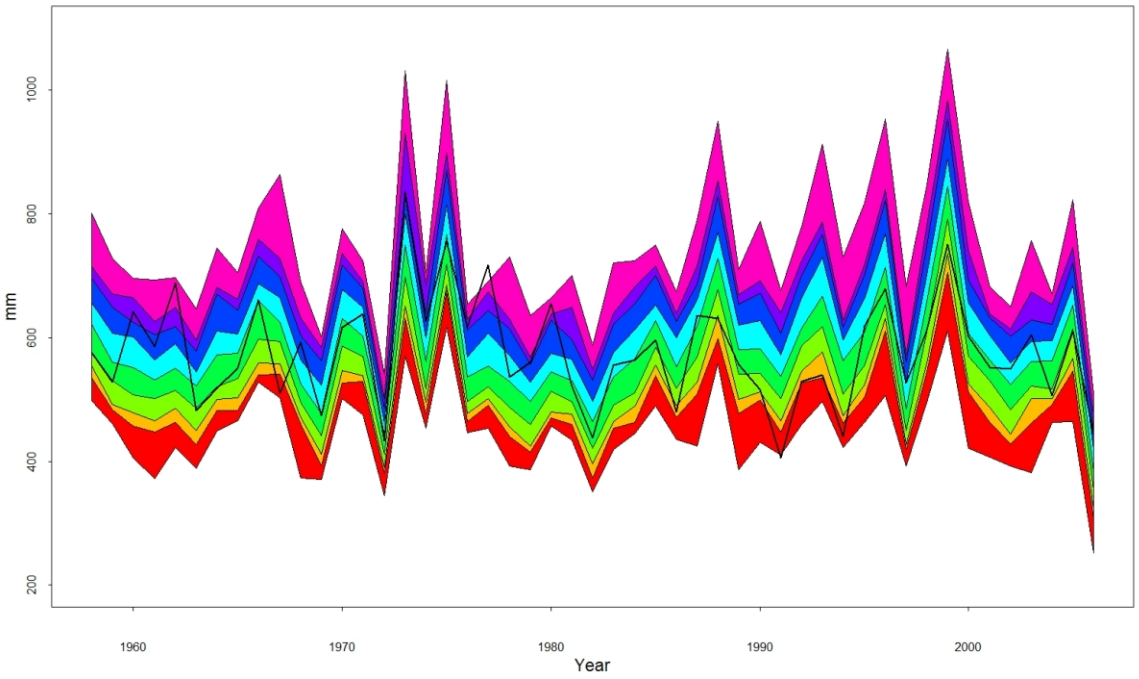


FIGURE 3.20: Probability Integral Transform for annual rainfall for the region: observed values as a sample from the simulated distribution.

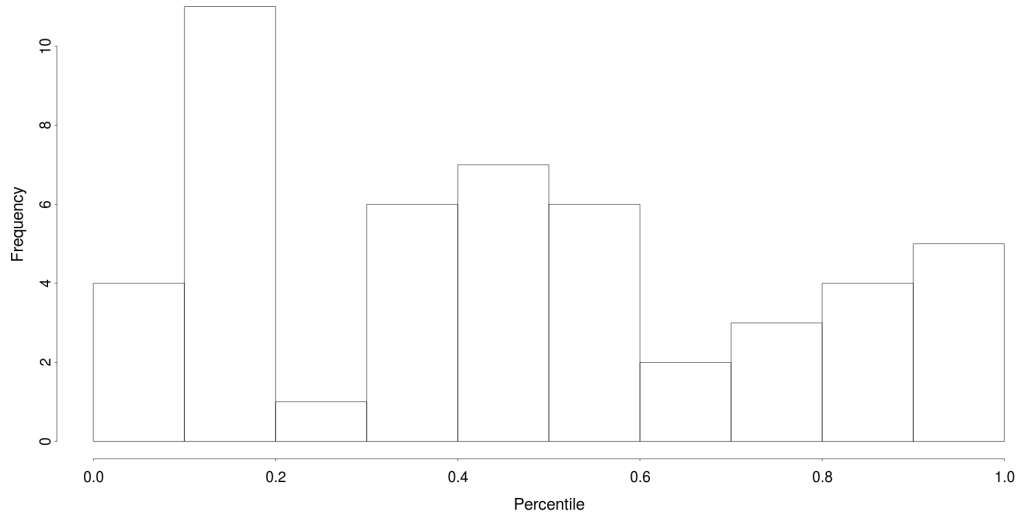


FIGURE 3.21: Observed time series (in black line) and rainbow range of annual regional average simulated distributions for 1970s validation period. Bands indicate the 0-5th, 5th-10th, 10th-25th, 25th-50th, 50th-75th, 75th-90th, 90th-95th and 95th-100th percentiles of the simulated distributions.

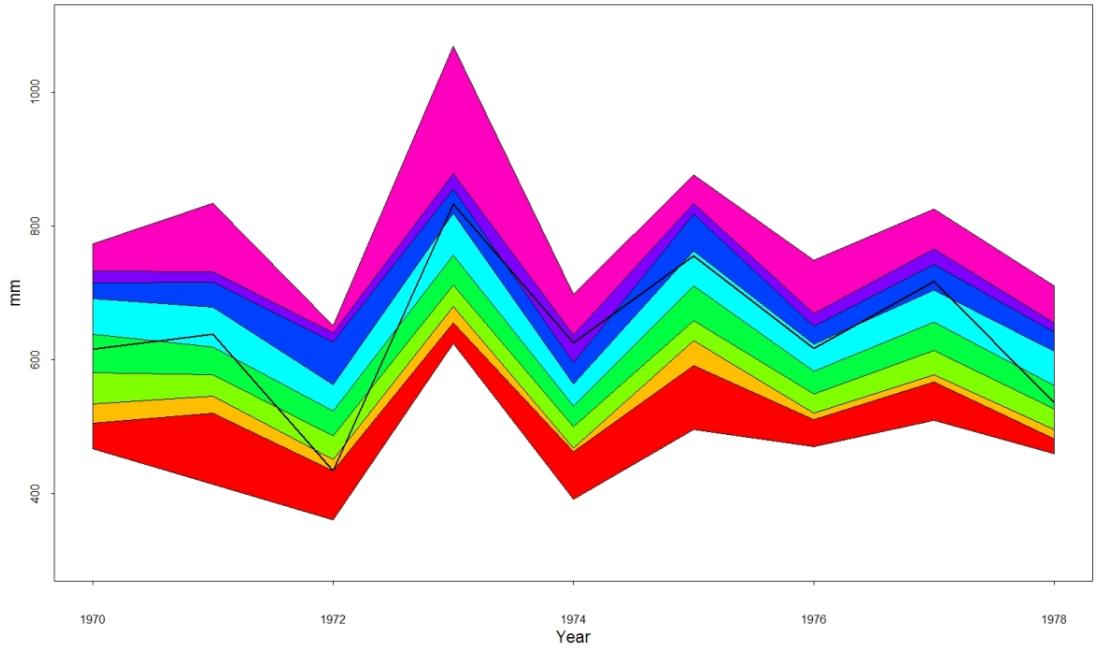


FIGURE 3.22: Observed time series (in black line) and rainbow range of annual regional average simulated distributions for 1980s validation period. Bands indicate the 0-5th, 5th-10th, 10th-25th, 25th-50th, 50th-75th, 75th-90th, 90th-95th and 95th-100th percentiles of the simulated distributions.

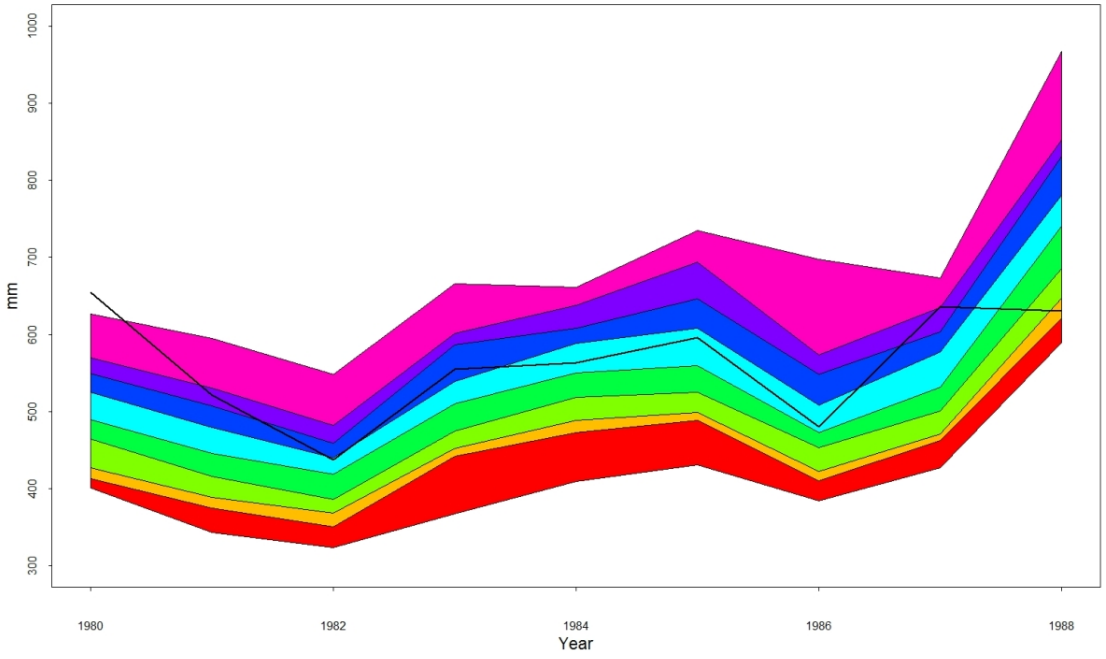
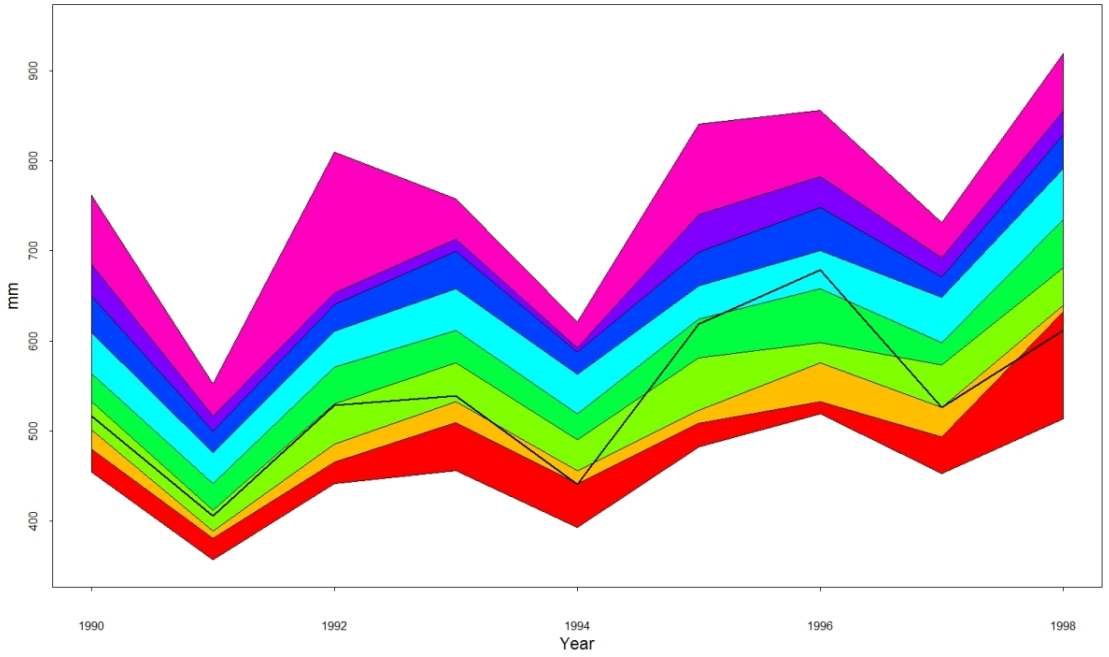


FIGURE 3.23: Observed time series (in black line) and rainbow range of annual regional average simulated distributions for 1990s validation period. Bands indicate the 0-5th, 5th-10th, 10th-25th, 25th-50th, 50th-75th, 75th-90th, 90th-95th and 95th-100th percentiles of the simulated distributions.



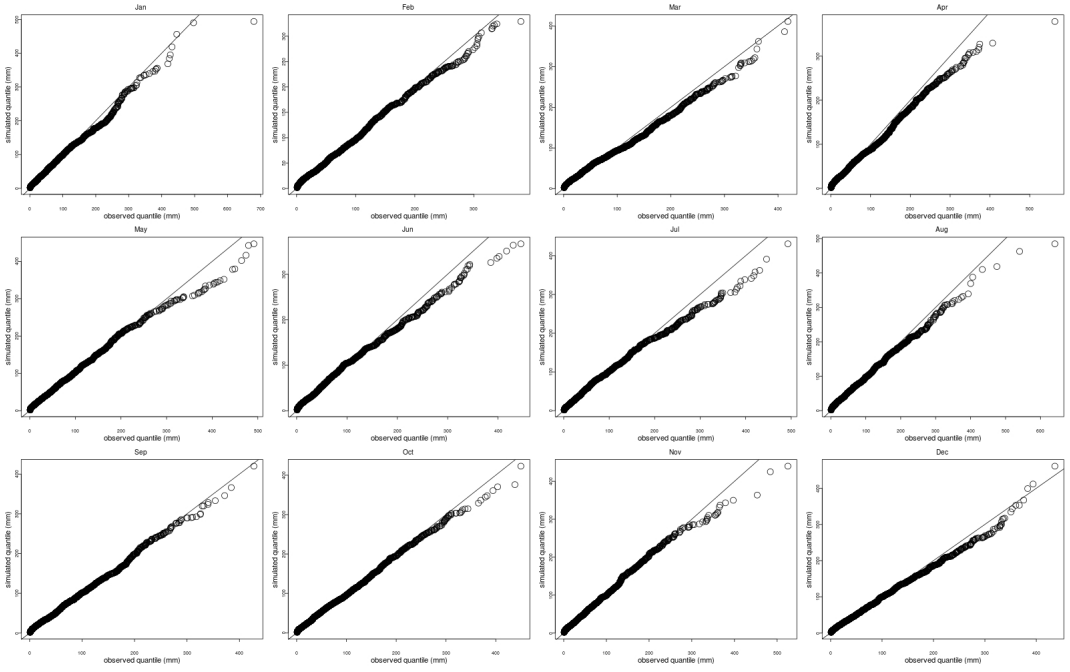
These plots show a good model performance overall: the precipitation structure is well captured by the simulated distribution over the calibration period as well as the three validation decades. In the calibration plot (Figure 3.19) the rainfall variability is well represented by the model and the observational gridded data fall within the simulation range. Figure 3.20 shows the Probability Integral Transform (PIT; Dawid, 1984) computed for the calibration model. The PIT can be used to assess the calibration of the modelled distribution and shows the number of years the observed total annual rainfall fell in between certain percentiles of the corresponding simulated distributions. If the simulations are realistic the observed values should look as if they have been sampled from the distributions. Based on the PIT the simulations appear to be 49 independent realizations from the uniform distribution (Leith, 2005b).

In the validation decades the model is able to reproduce the year to year rainfall except for extremely dry years. In the first validation decade (Figure 3.21) the observed rainfall tends to fall within the upper half of the simulated distributions, with a tendency for

the model to underestimate the areal average rainfall value. The opposite is true for the 1990s (Figure 3.23). Such bias in the simulation is very likely related to the decadal variability unexplained by the model (Figure 3.16). However, in general the variability is very well reproduced throughout the three simulated decades.

For each month of the year, quantile-quantile plots of observed versus simulated monthly rainfall have been produced for each decade of simulation (example for the 1990s in Figure 3.24). An overall agreement is visible among observed and simulated rainfall distributions, although there are problems in the upper tail of the plot, where the simulations tend to underestimate highest precipitation values. However this does not appear to be a problem in the current project since the final aim of the work is the investigation of water shortage during the crop growing season so that the primary interest is in the lower tail of the distribution.

FIGURE 3.24: Quantile-quantile plot for observed vs simulated monthly rainfall distribution for the 1990s validation period.



3.5.1 Applicability at local scales

The models developed so far have been showed to perform reasonably in reproducing rainfall variability at subcontinental scale. However, impact studies often require local-scale information. Hence, a region within the subcontinent was considered to investigate the applicability of the developed models to spatially downscale large-scale climate drivers to deduce rainfall information at narrower spatial scale. The region extrapolated is the Limpopo Province – Greater Sekhukhune district in South Africa (Figure 3.25; and following case study presented in Chapter 5). As for the subcontinental scale results in Figures 3.19 – 3.23, for this local region the simulated distributions provide a good representation of interannual variability but are unable to capture the decadal-scale fluctuations. On the other hand, the quantile-quantile plots for the three decades displayed extremely poor results (not shown here), indicating that the model is not suitable for generating monthly rainfall at the relatively local scales typically required for impact study purposes. Reasons for this may lie in assumptions made during the modelling process and inherent characteristics of the region such as:

a) the use of a common shape parameter of the gamma distribution function at all spatial and temporal location (Section 3.3.1);

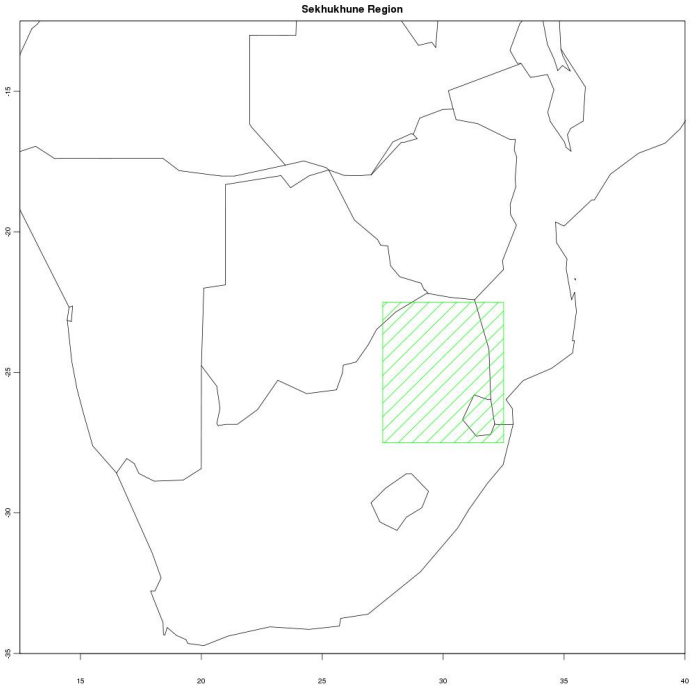
b) climate and topographical heterogeneity in the region may not allow for the development of a statistical model able to precisely represent monthly rainfall at each spatial location;

c) inhomogeneities induced by the gridding procedure used for the creation of the gridded monthly rainfall dataset may contribute to the low performance of the simulations by the statistical model at local scale.

3.6 Conclusion

The work presented in this chapter aimed to characterise southern African rainfall variability and gain deeper knowledge of its relationships with other aspects of the climate system considered simultaneously. The potential of the chosen statistical framework for

FIGURE 3.25: Sekhukhune case study region. The green shading delimits the four grid cells enclosing the area of interest.



simulating synthetic rainfall sequences was also tested.

Such comprehensive investigation of the climate system provides an understanding of how climate indices act in combination, as well as how the effect of one factor may itself be modulated by others. The use of GLMs allows to represent the behaviour of the entire system using a statistical model, which gives a more realistic assessment of uncertainty than simply working with anomalies, as done in most previous statistical studies of the climate system. GLMs also allow the selection of a smaller number of covariates than building a regression model for every single grid cell.

Out of almost twenty atmospheric and oceanic features reported in the literature as potentially significant, only half a dozen proved to be key drivers in explaining part of the variability in southern African monthly rainfall. The ability to isolate key drivers is another advantage of studying the various climate indices simultaneously, since this provides the opportunity to discriminate between competing explanations of the observed relationships. This illustrates the potential for analyses of single drivers to yield an incomplete understanding of the climate system.

The modelled effects of the main climate indices upon precipitation occurrence and

amounts were broadly in agreement with what is already known about the climatic drivers of southern Africa rainfall. The isolation of the jet stream effect in explaining part of the rainfall variability in the considered region is new, however. On the other hand, where a statistical correlation between rainfall variability and climate index was not detected here, the current study does not necessarily contradict the previous research reported in the literature since most of the climate indices did indeed appear significant in our analysis when considered separately. Thus, this work shows that after accounting for the most significant climate indices, any additional effect of the remainder is negligible after a stringent criteria of 1% significance was considered.

Due to the complex orography of the region, many model terms were required to represent the regional variability across the subcontinent. The main time-varying factors affecting precipitation upon the studied region were firstly relative humidity followed by El Niño and mean southern hemisphere temperature.

Interestingly, diagnostics revealed an interdecadal variability in rainfall amounts that could not be accounted for by any of the drivers considered.

The analysis showed the ability of the model to simulate gridded monthly rainfall data sets at a subcontinental scale. Since one of the aims of the current research is the investigation of future precipitation scenarios for the subcontinent, the relationships revealed by the analysis reported here have been used to downscale outputs from general circulation models at subcontinental scale (Chapter 4).

At the scale of a few grid cells, however, the simulation performance of the developed models was inadequate. Therefore the statistical model, while useful to illustrate the southern African rainfall climate drivers, was not considered adequate for impact study purposes and a specifically local scale model was developed later using the same methodology (Chapter 5).

Chapter 4

Future regional Rainfall Projections

4.1 Introduction

In the previous chapter a model relating subcontinental rainfall variability to large-scale atmospheric and climate drivers was developed. In this chapter the extent to which this model can be used to remedy some of the deficiencies of GCMs for the simulation of precipitation over southern Africa is explored. As demonstrated previously, the statistical model provides credible simulations only at regional scale; therefore this is the focus of this chapter.

In the next sections large scale climate predictors, from GCMs, have been linked to southern African precipitation, considered as the local predictand, using the same models presented in Chapter 3. In such way, outputs from a set of GCMs have been downscaled and multiple synthetic precipitation sequences randomly generated without using climate models' precipitation directly. The obtained envelope of simulated rainfall may be interpreted as representing the uncertainty due to internal variability while simulating the climate response to a given emissions scenario for a given climate model and initial condition. Those projected envelopes of future monthly precipitation provide probabilistic information about what may be expected in the studied region.

Section 4.2 illustrates the results for a 20st century control period, where results of simulated precipitation have been compared with the observed record and GCM projections. This exercise was necessary in order to test the statistical model ability to simulate 'present climate', and gave useful insight into the GCM biases in representing southern African averaged precipitation. In Section 4.3 results for a forty-year period in the 21st century are presented in order to illustrate future precipitation projections for the region.

4.2 20th century control period

4.2.1 Data used

As in Chapter 2, the datasets used were derived from the WCRP-CMIP3 multi-model dataset. The considered climate models are listed in Table 4.1 and are a subset of the GCMs used in Chapter 2 due to the availability of data required as climate covariates. As already stated in Section 1.3.1, the climate models do not provide a 'true' representation of the climate system, but simplification of the same due to the computational expense and limited understanding of such a complex system. The models use a range of different schemes in modelling the interaction within the climate system as well as parametrizations. Consequently they may produce different projections about what will occur in the future. The incorporation of outputs for all available climate models provides a means of exploring uncertainty due to the choice of GCM.

A 41-year period, from 1957 – 1998, was investigated. A single output run was randomly selected among the available ensemble members (see Section 2.2) for each climate model variable required as a covariate by the statistical models. The choice of one single run for each climate model has been considered enough in the investigation of the multi-model uncertainty span. Indeed, according to the Figure 2.3 and Northrop (2010), the highest source of variability in both the 20th and 21st centuries precipitation means, is due to the choice of GCM and only subsequently the run.

Each climate model field was reinterpolated to the CRU TS3 grid (2.5°× around 2.5°) for consistency with the analysis reported in the previous chapters. Once more the chosen domain has been defined as the area lying between 12° – 40°S and 0° – 42°E.

The monthly climate indices required by the precipitation occurrence and amounts models were then calculated for each GCM, as follows:

- El Niño 3.4 SST index
- SHT mean index
- SAM SLPs index
- SWIO SSTs index
- Jet stream intensity and relative position effect indices
- RHUM index

The first five of these indices were created by extracting the corresponding grid cells of the atmospheric or oceanic field, after reinterpolation, and computing the index in the same way as described in Section 3.2.
For the RHUM index, the temporal scores were computed both from the first loading pattern obtained from a principal component analysis computed individually for each GCM, and using the loading pattern obtained from the NCEP reanalysis relative humidity field (see Figure 3.4). The second option was later dismissed. Since some GCMs appeared to have relative humidity loading patterns highly different from what detected in the NCEP reanalysis field, the use of such relative humidity index induced problems during the rainfall simulations process, such as the generation of unreasonable rainfall amounts.

4.2.2 Simulations

Following the methodology described in Section 3.3.7, the precipitation occurrence and amounts models were used jointly to simulate synthetic rainfall sequences conditioned on the outputs of each of the GCMs. The simulations were initialized with historical data for December 1957 and run for 41 years, until December 1998. 100 precipitation sets were simulated for each climate model and yearly and monthly rainfall summary

Centre	*Country*	*Model Name*
Bjerknes Centre for Climate Research	Norway	BCCR-BCM2.0
Canadian Centre for Climate Modelling & Analysis	Canada	CGCM3.1(T47)
Canadian Centre for Climate Modelling & Analysis	Canada	CGCM3.1(T63)
Météo-France / Centre National de Recherches Météorologiques	France	CNRM-CM3
Commonwealth Scientific and Industrial Research Organisation	Australia	CSIRO-Mk3.0
Commonwealth Scientific and Industrial Research Organisation	Australia	CSIRO-Mk3.0
Max Planck Institute for Meteorology	Germany	ECHAM5/MPI-OM
US Dept. of Commerce / NOAA / Geophysical Fluid Dynamics Laboratory	U.S.A.	GFDL-CM2.0
US Dept. of Commerce / NOAA / Geophysical Fluid Dynamics Laboratory	U.S.A.	GFDL-CM2.1
NASA / Goddard Institute for Space Studies	U.S.A.	GISS-EH
NASA / Goddard Institute for Space Studies	U.S.A.	GISS-ER
LASG / Institute of Atmospheric Physics	China	FGOALS-g1.0
Institute for Numerical Mathematics	Russia	INM-CM3.0
Institut Pierre Simon Laplace	France	IPSL-CM4
Meteorological Research Institute	Japan	MRI-CGCM2.3.2
National Center for Atmospheric Research	U.S.A.	PCM
Hadley Centre for Climate Prediction and Research / Met Office	UK	UKMO-HadCM3
Hadley Centre for Climate Prediction and Research / Met Office	UK	UKMO-HadGEM1

TABLE 4.1: Climate models considered in the statistical downscaling of GCM outputs.

statistics were computed.

Initially, the simulations for some GCMs generated some unrealistically high values (in excess of 4000 mm/month in the south of the region). These were subsequently identified as associated with the jet stream indices as computed from the GCM outputs, which proved to be problematic in this context. Specifically, in the southernmost grid cells (33.75°S and 31.20°S) the jet stream covariates led to unstable rainfall simulation in late winter months (from August throughout October). During those months the projected subtropical jet stream was far too north, causing the POSEFF index, defined in Section 3.4.2 and equation 3.5, to take unrealistically high values. The final result was the simulation of unreasonably high rainfall amounts for a number of GCMs. An attempt was made to overcome the problem by scaling the POSEFF covariate to the same mean and variance as the NCEP values. This failed to resolve the problem however. Therefore, it

was decided to remove all the covariates concerning the jet stream and refit the models for this exercise.

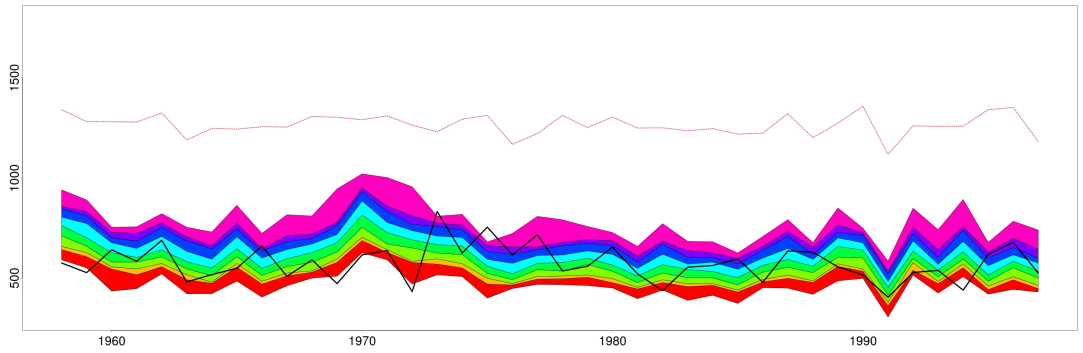
Figure 4.1 shows the simulated distribution of regionally-averaged annual precipitation totals, driven by each of the climate model outputs listed in Table 4.1. The colour scale shows the quantiles of these distributions as in Chapter 3. The plotted years are hydrological, starting from July to June with the year representing the time of the start of the rainy season. The black thick solid line represents the observed averaged precipitation and the red dotted thin line the precipitation obtained directly from the corresponding climate model.

The results indicate good agreement between the simulated synthetic rainfall sequences and the observed time series, as in most of the cases the range of the simulations encompass the observed time series, indicating the ability of the statistical models to represent the detected rainfall variability. Overall the statistical models give better results than the corresponding climate models in representing total average yearly precipitation over the region. For example, ten of the climate models systematically over project the subcontinental precipitation throughout the studied period, by as much as $200 - 300$ mm/year (Figure 4.1, plots e, h, i, k, p, q and r), 400 mm/year (Figure 4.1, plot j) and $600 - 700$ mm/year (Figure 4.1, plots a and d). Only in one case, see Figure 4.1 (m) does the projected rainfall by the climate model appears to be systematically lower than that recorded. Such biases in the representation of tropical precipitation by GCMs are already known. In particular, most of the climate models produce a pattern with excessive precipitation off the equator but insufficient precipitation on the equator, often associated with an excessive and overly narrow SST cold tongue that extends too far west into the western Pacific (double ITCZ mechanism; Biasutti et al. 2006; Lin 2007). Interestingly the interannual variation in the statistically simulated rainfall series does not always align with that in the rainfall projected by the GCMs. This suggest that the parametrization in the GCMs may not respect some of the identified relationships between predictand and predictor.

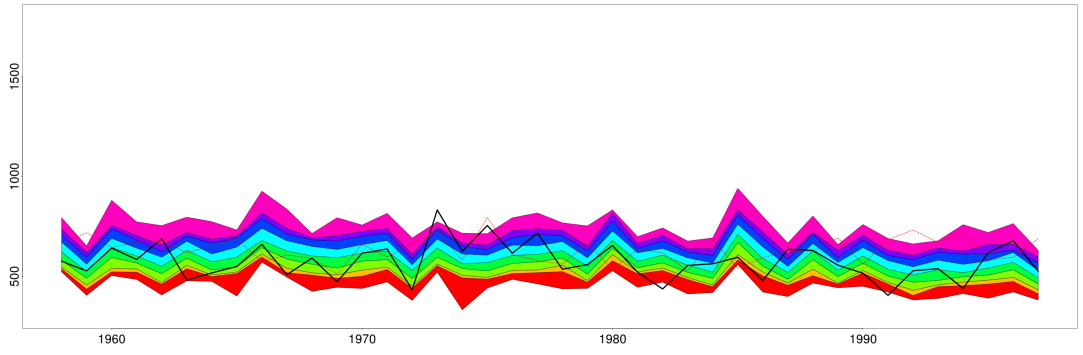
The results of a mean seasonal cycle averaged across the region and period are presented in Figure 4.2. In general the statistical model is able to reproduce quite well the seasonal cycle of the region in question. On the contrary, the climate models seem to be unable to represent the seasonality correctly, with the majority of them overpredicting

precipitation amounts all year round (see for example Figure 4.2, plots a, d, e, g, h, i, j, k, p, q and r) and a few overpredicting during the dry season and underpredicting during the wet (Figure 4.2, plots b, c and m).

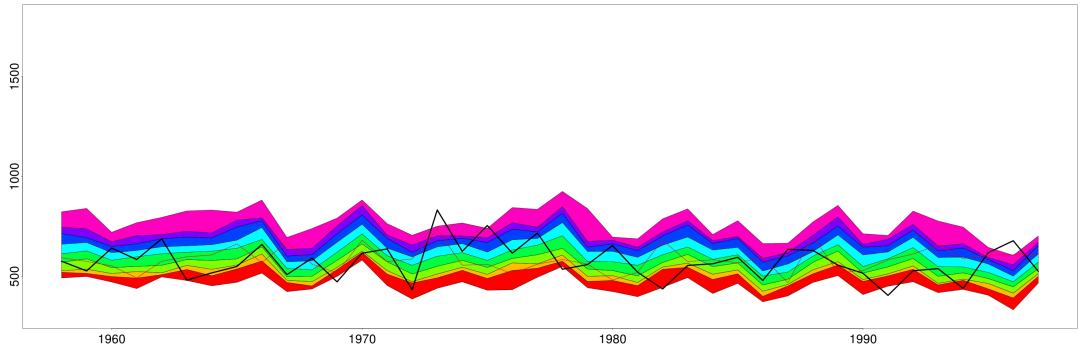
FIGURE 4.1: Rainbow range of simulated distribution, observed time series (in black thick solid line) and climate model projection (in dashed thin red line) of annual southern African precipitation averaged over the entire region for the control period and eighteen GCMs.



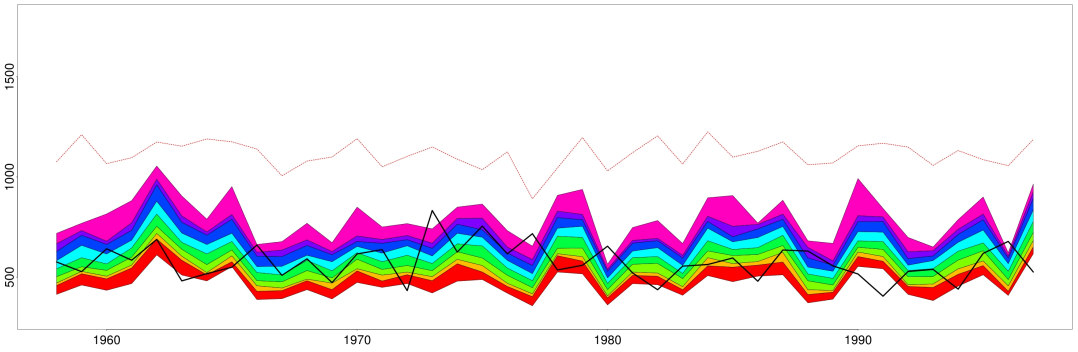
(a) BCCR-BCM2.0



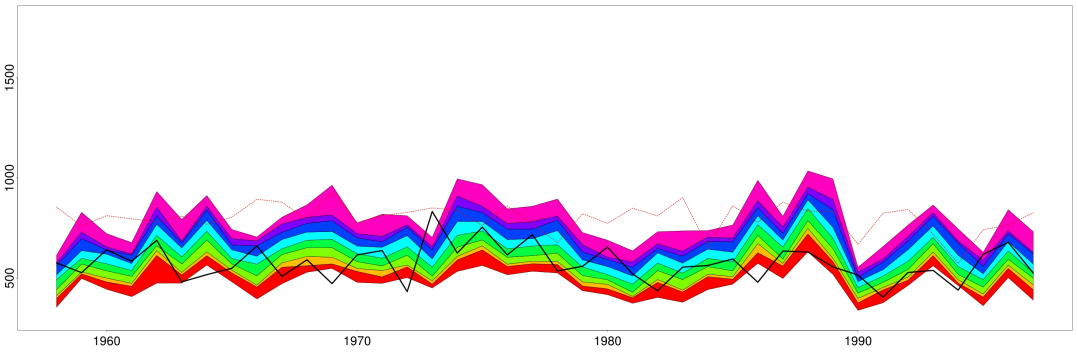
(b) CGCM3.1(T63)



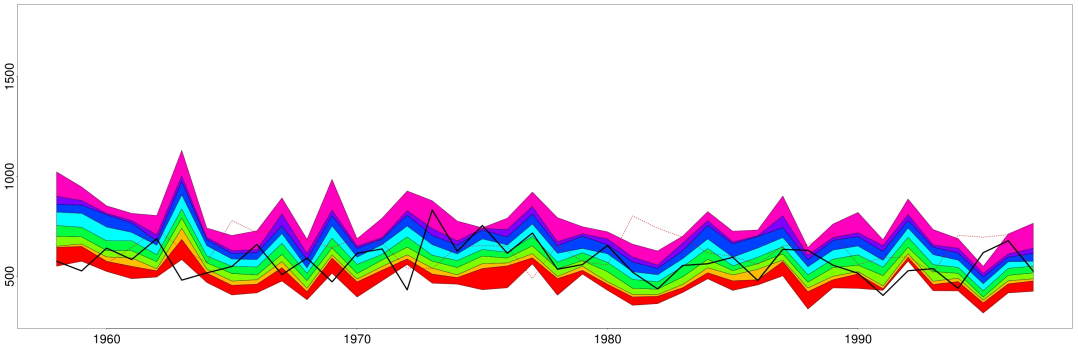
(c) CGCM3.1(T47)



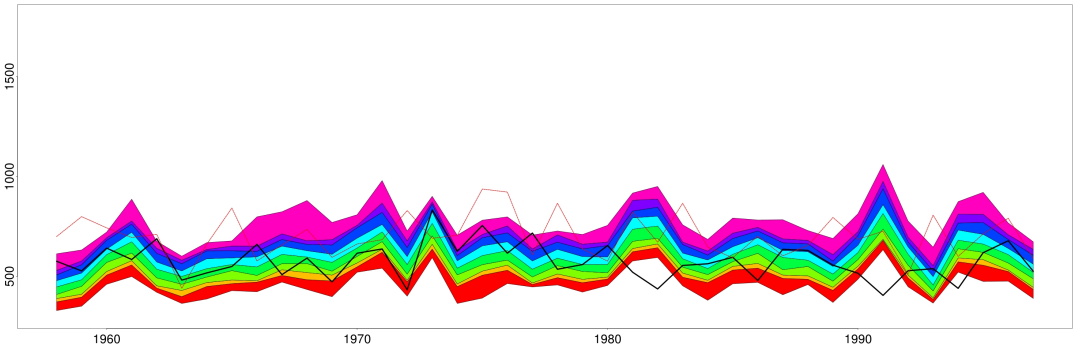
(d) CNRM-CM3



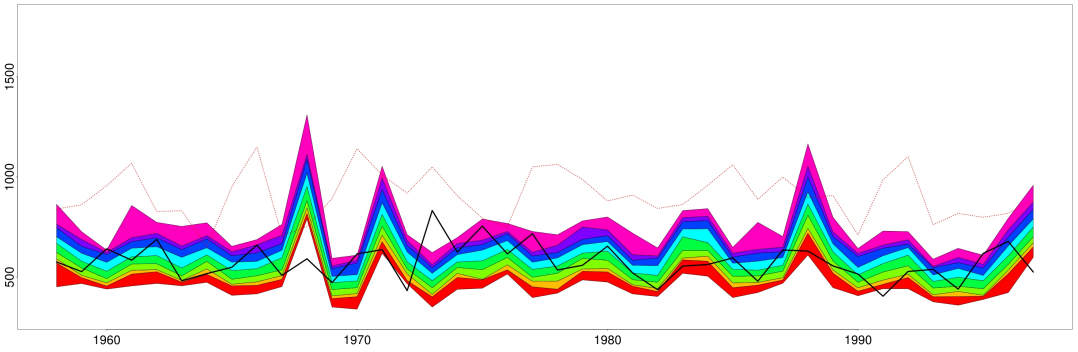
(e) CSIRO-Mk3.0



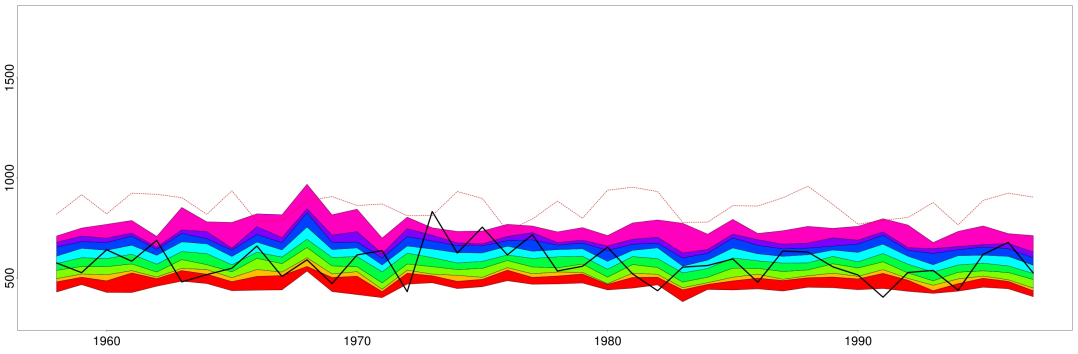
(f) CSIRO-Mk3.5



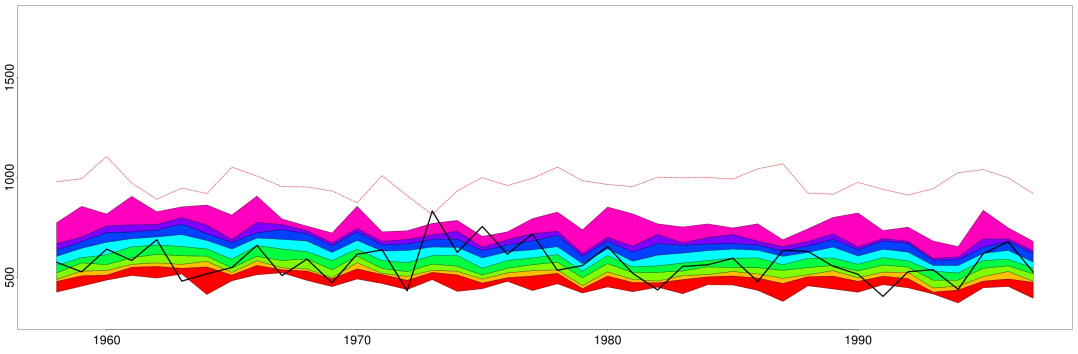
(g) GFDL-CM2.0



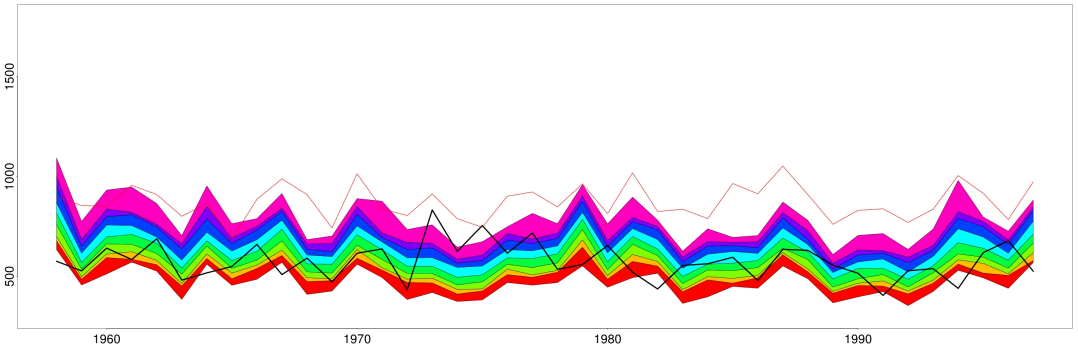
(h) GFDL-CM2.1



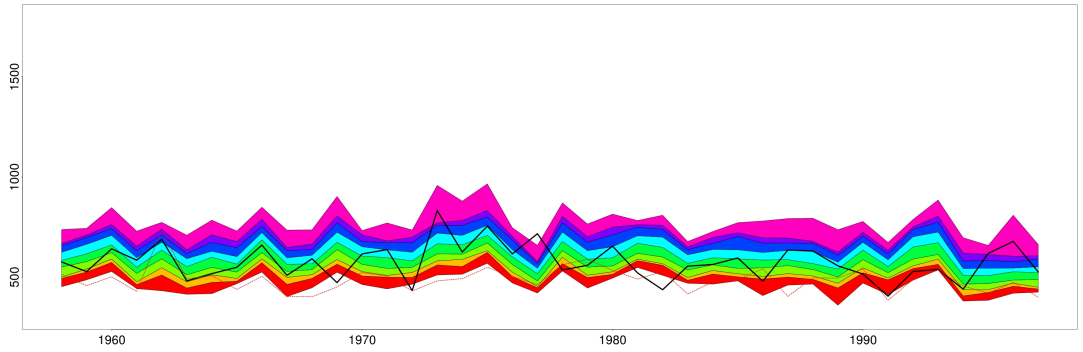
(i) GISS-EH



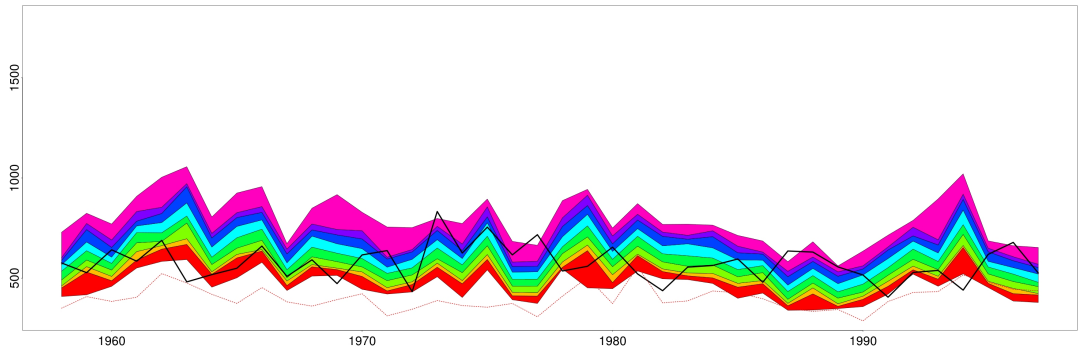
(j) GISS-ER



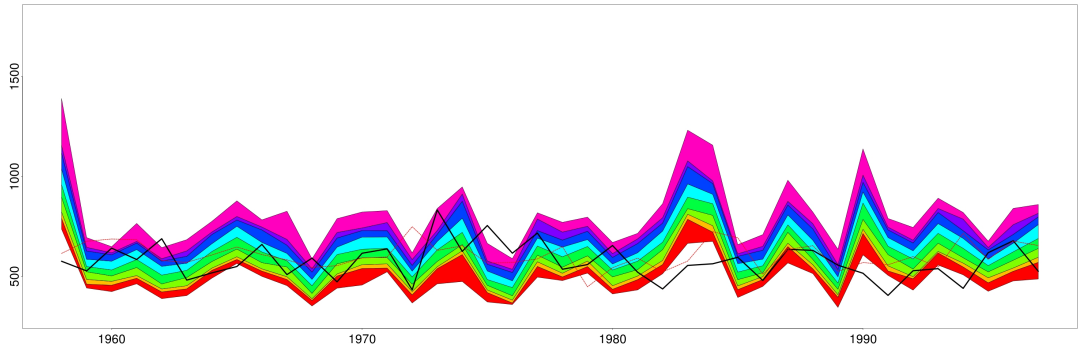
(k) IAP-FGOALS



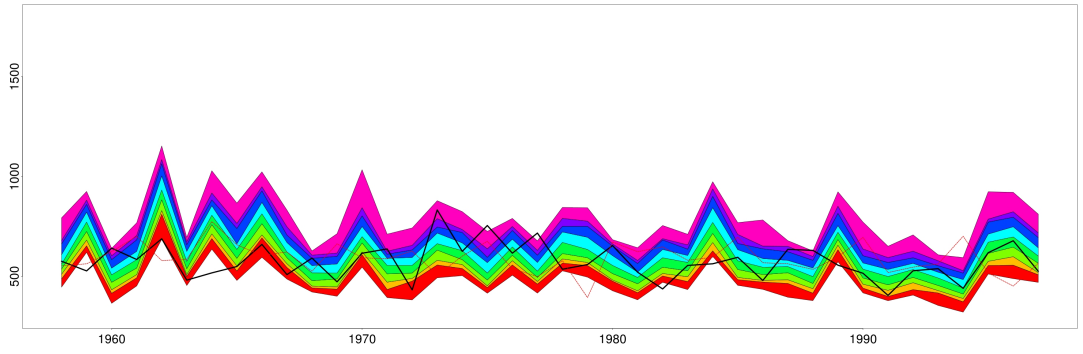
(l) INM-CM3.0



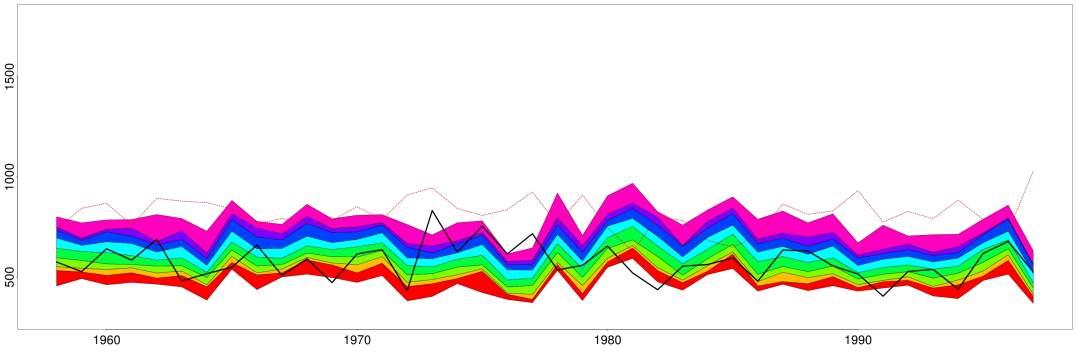
(m) IPSL



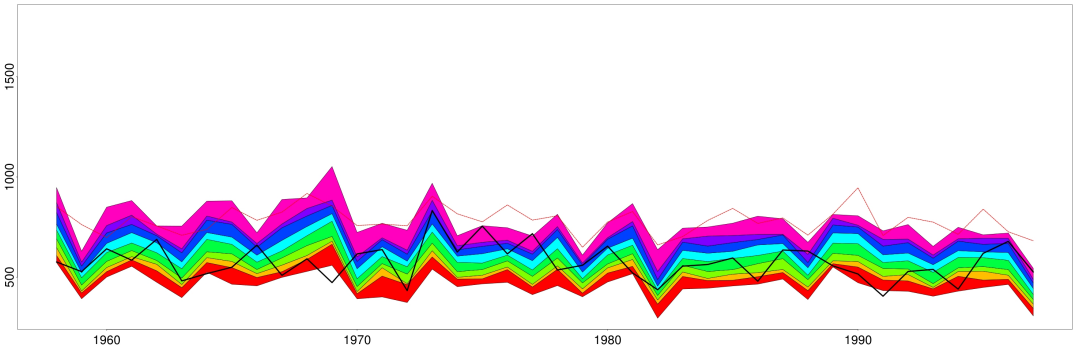
(n) ECHAM5



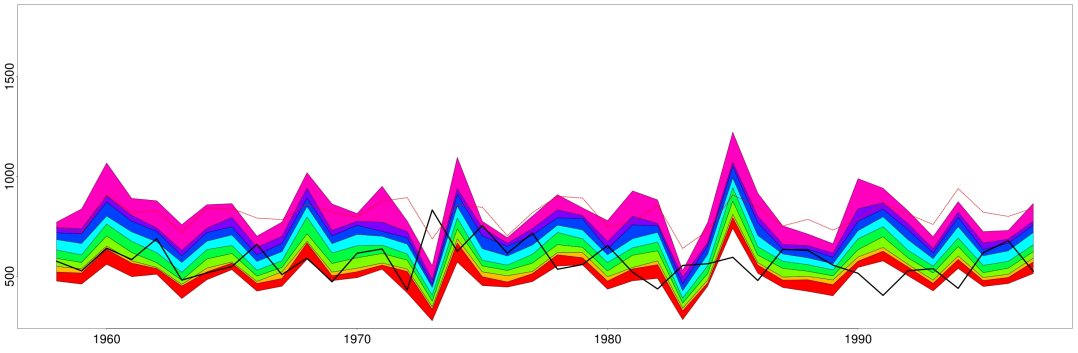
(o) MRI-CGCM



(p) PCM

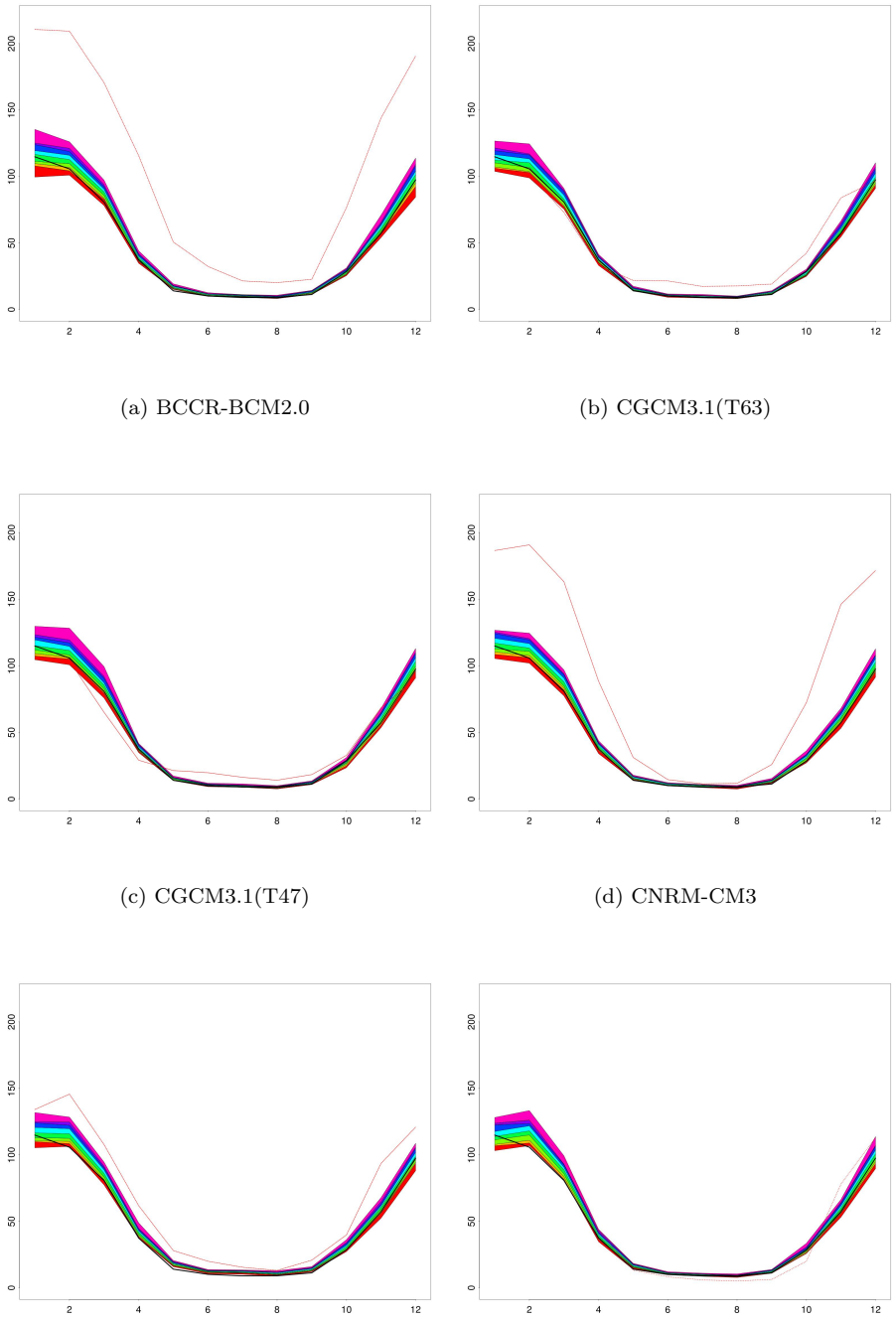


(q) UKMO-HadCM3



(r) UKMO-HadGEM1

FIGURE 4.2: Rainbow range of simulated monthly climatology, observed climatology (in black thick solid line) and GCMs derived climatology (in dashed thin red line) of southern African precipitation averaged over the entire region for the control period and eighteen GCMs.



(a) BCCR-BCM2.0

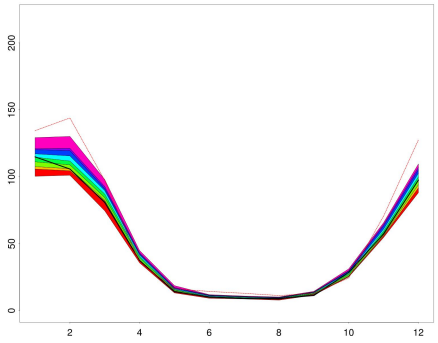
(b) CGCM3.1(T63)

(c) CGCM3.1(T47)

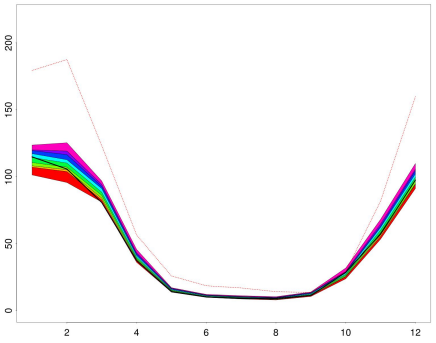
(d) CNRM-CM3

(e) CSIRO-Mk3.0

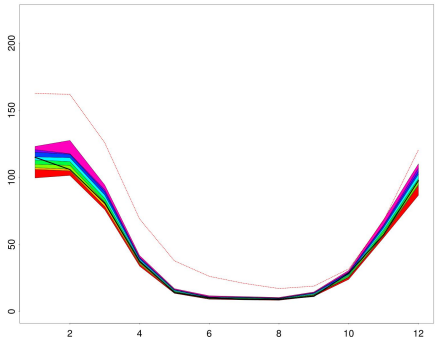
(f) CSIRO-Mk3.5



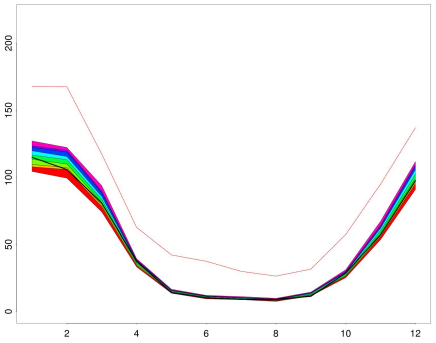
(g) GFDL-CM2.0



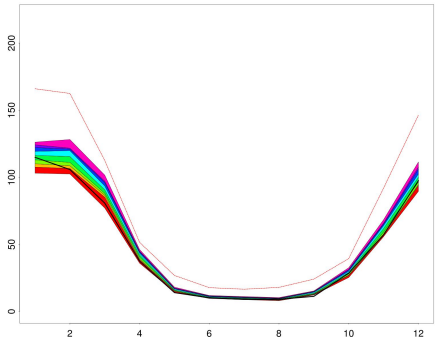
(h) GFDL-CM2.1



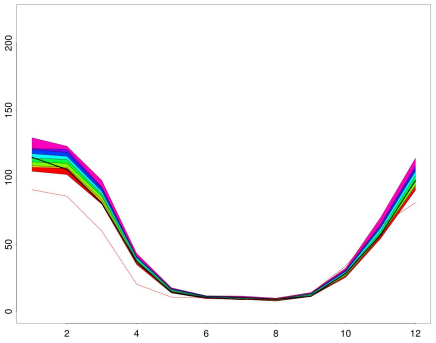
(i) GISS-EH



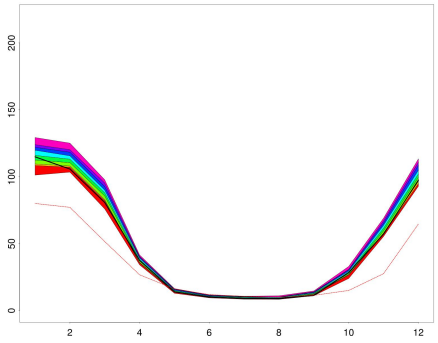
(j) GISS-ER



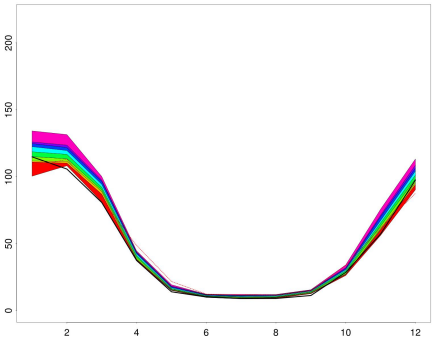
(k) IAP-FGOALS



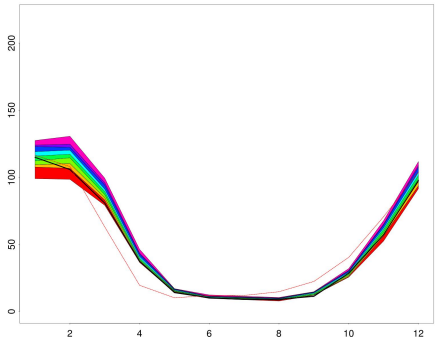
(l) INM-CM3.0



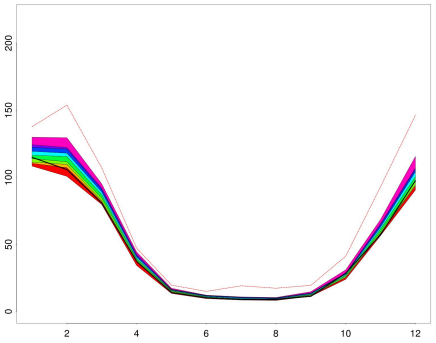
(m) IPSL



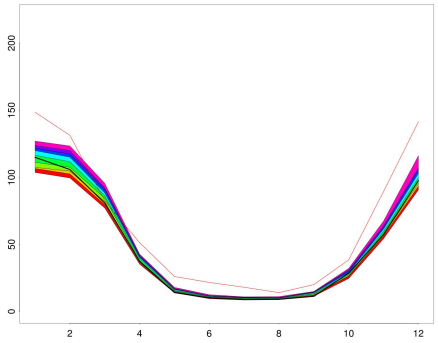
(n) ECHAM5



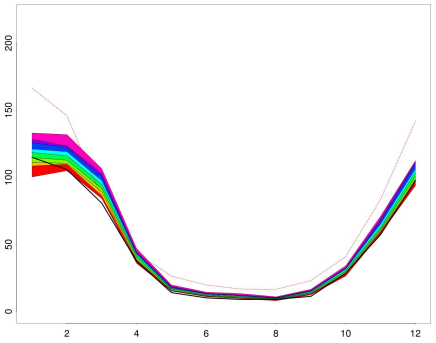
(o) MRI-CGCM



(p) PCM



(q) UKMO-HadCM3



(r) UKMO-HadGEM1

4.3 Future rainfall projection under a climate change scenario

4.3.1 Data used

Similarly to the work presented in the previous section, the statistical models driven by GCM outputs were used to gain insight on the precipitation variability over the subcontinent, this time for a selected period in the future. A forty-year interval, from 2057 – 2098, was considered.

Out of the plausible future climate scenarios constructed by the IPCC for the investigation of the potential consequences of anthropogenic climate change (see section 1.3), the a1b scenario was chosen as an example for the study of the future climate. As defined in the Special Report (IPCC, 2000), the future economic system under the a1b scenario is projected to rapidly grow using a balance in energy sources with a peak in population in the middle of the 21st century.

This exercise used the same eighteen climate models as in the previous section (Table 4.1). Similarly to the 20th century study, a single ensemble member for each variable has been randomly selected among the available ones and the climate model outputs have been processed as described in Section 4.2.

4.3.2 Simulations

The simulation results have again been averaged annually and seasonally for the entire region. For uniformity in the results and simplicity in their interpretation the y-axis scale has been kept identical to the one used in Figures 4.1 and 4.2 and average values for the corresponding 20th century simulations added to the plots (black thick lines).

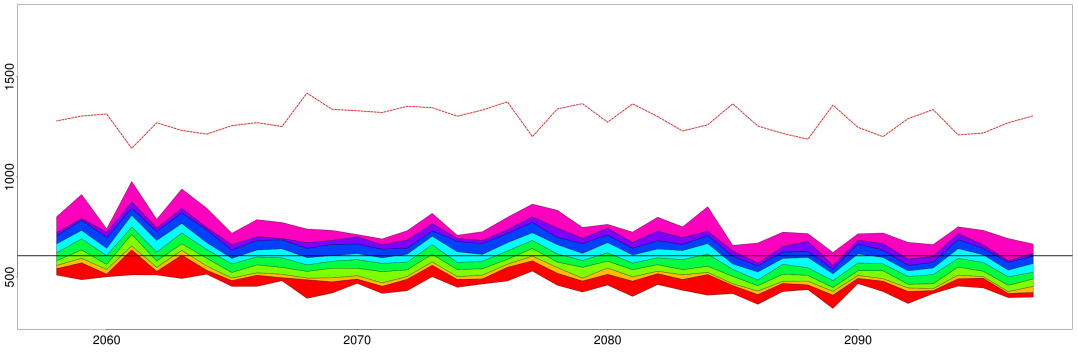
Figure 4.3 shows the quantiles of 100 simulations for the period 2058 – 2097 conditional on each GCMs' future projections. The thick black lines represent the overall mean of the 20th century simulations. Generally the precipitation range of the simulation is similar to that obtained in the 20th century simulations, although for some of the GCM driven simulations a tendency to produce drier annual precipitation is visible especially in the final two decades of the 21st century (Figure 4.3, plots a, d, e, f, i, j, l and o).

In some of the plots (Figure 4.3, plots c, e, f and o) a negative trend in the simulations is visible, with an overall decrease in precipitation during the investigated forty-year period of as much as 200 mm/year. In general the year-to-year variability in the simulations driven by the GCMs derived climate drivers and the span of the simulation seems consistent in the two studied periods, with some GCM driven sequences still experiencing higher variability (Figure 4.3, plots g, h, n and r) than others (Figure 4.3, plots c, i and j).

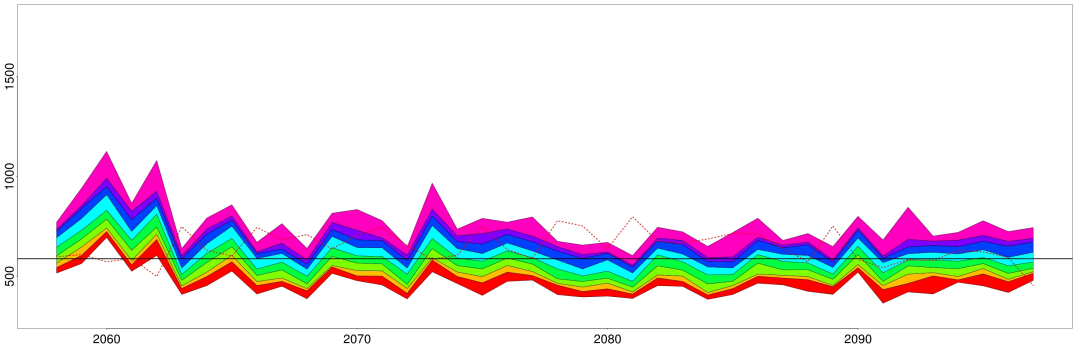
When comparing the rainfall obtained directly from the GCMs with the statistically simulated distributions, the former tend to be higher (Figure 4.3, plots a, d, h, i, j, k, p and r) or of similar magnitude (Figure 4.3, plots b, c, e, f, g, n, o and q). Only in a few cases does the climate models' rainfall appear drier than the simulated (Figure 4.3, plots l and m). This is essentially the same as for the control period results reported above.

The simulations of the average seasonal cycle in the future 40 years of investigation are shown in Figure 4.4. During the rainy season the GCMs' own monthly precipitation means tend to be higher than the simulations, especially from December to February (Figure 4.4, plots a, d, i, j, k and p). For some of the GCMs, projected precipitation is lower than the corresponding simulated monthly means during transition months: March, April, May, September and October (Figure 4.4, plots f, h, o, p, q and r). On the other hand, the range of statistically simulated rainfall appears narrower at the peak of the rainy season (December-February) suggesting a reduction in the uncertainty under future climate conditions. However, compared with the average seasonal cycle obtained from 20^{th} century simulations (thick black lines) there is no clear signal of change in the simulated distributions for the 21^{st} century.

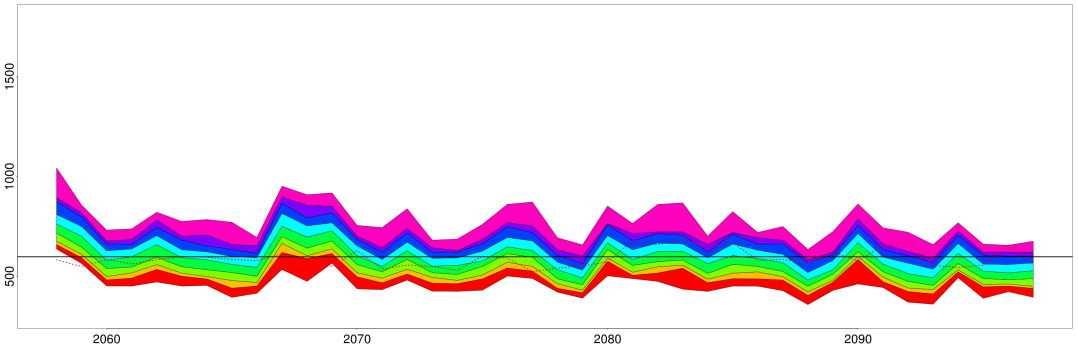
FIGURE 4.3: Rainbow range of simulated distribution and climate model projection (in dashed thin red line) of annual southern African precipitation averaged over the entire region for the future period and eighteen GCMs. Dashed thick black line indicating overall mean for the corresponding 20^{th} century simulation.



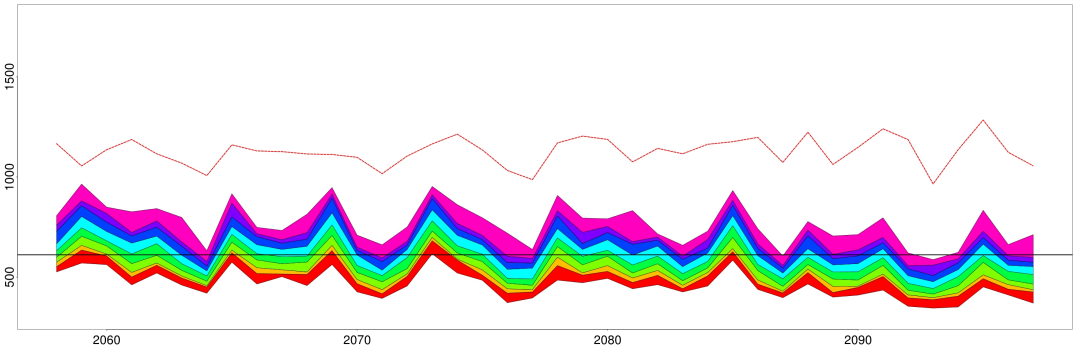
(a) BCCR-BCM2.0



(b) CGCM3.1(T63)

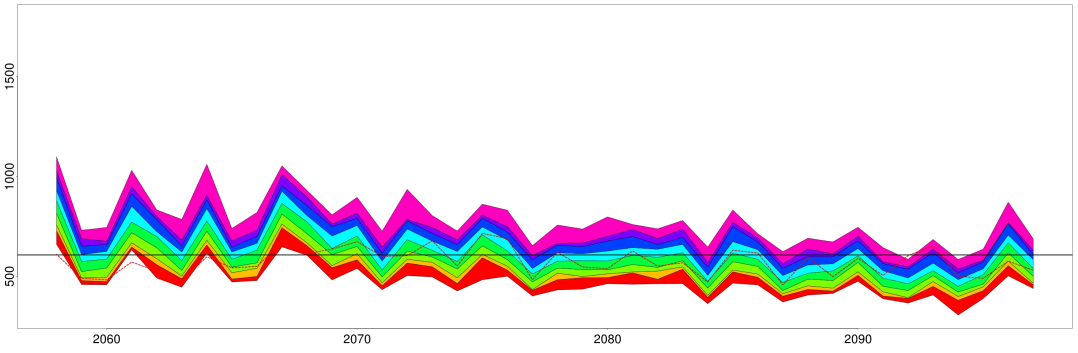


(c) CGCM3.1(T47)

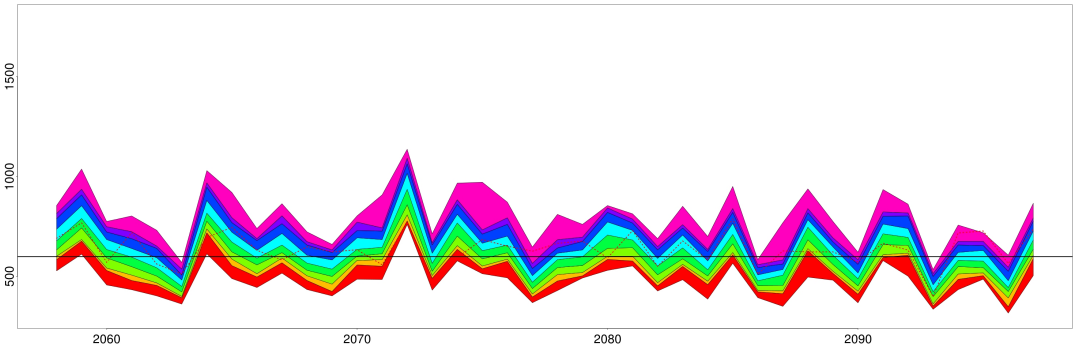


(d) CNRM-CM3

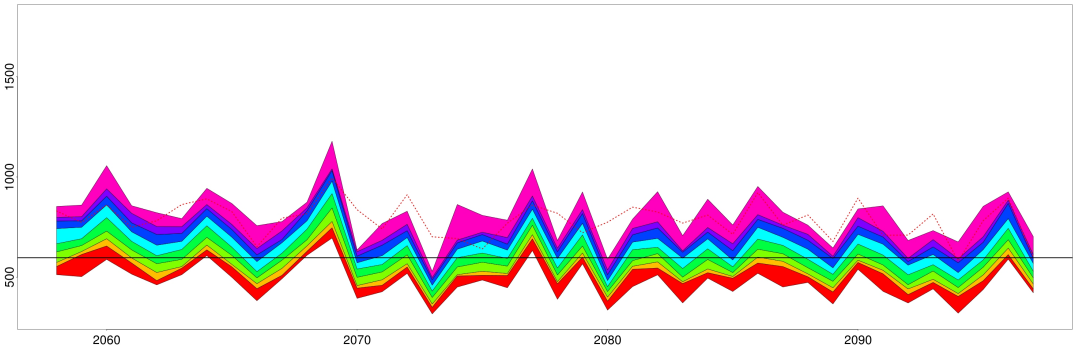
(e) CSIRO-Mk3.0



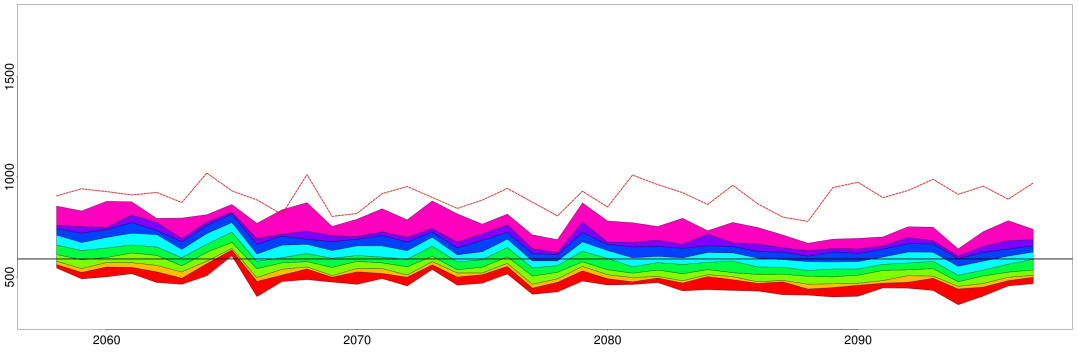
(f) CSIRO-Mk3.5



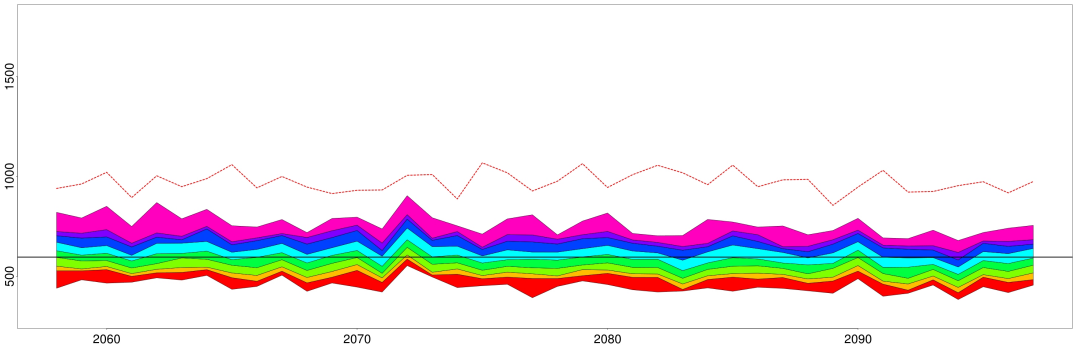
(g) GFDL-CM2.0



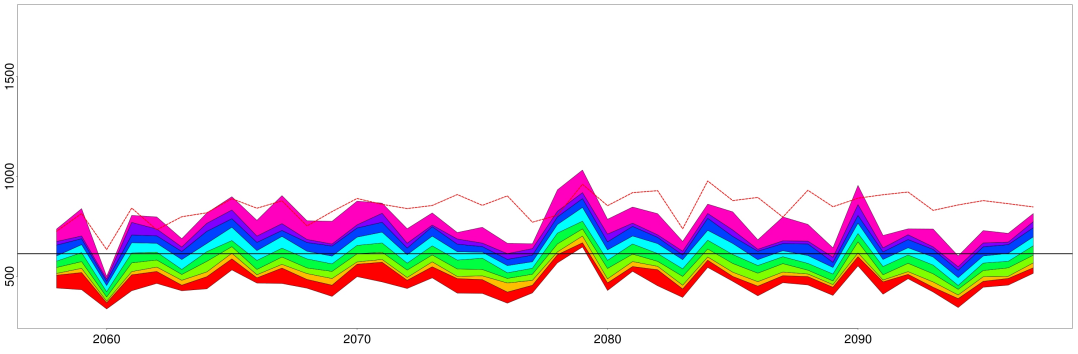
(h) GFDL-CM2.1



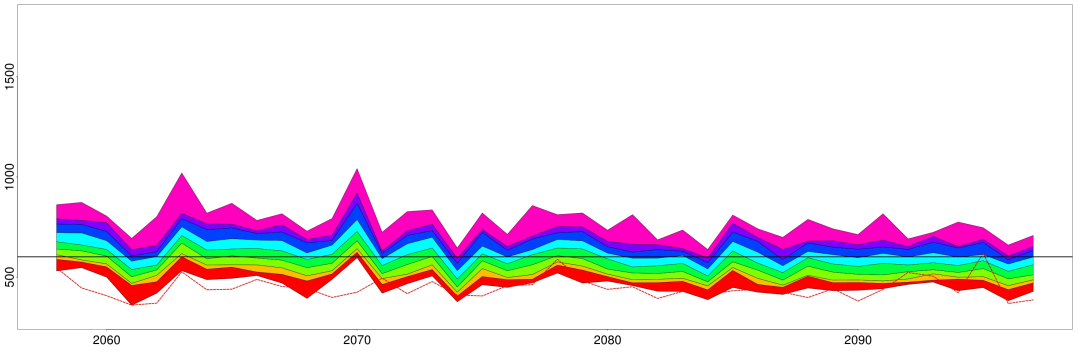
(i) GISS-EH



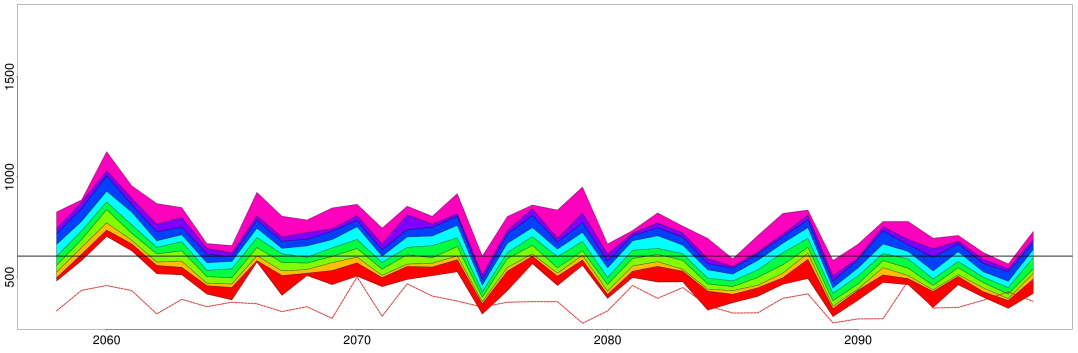
(j) GISS-ER



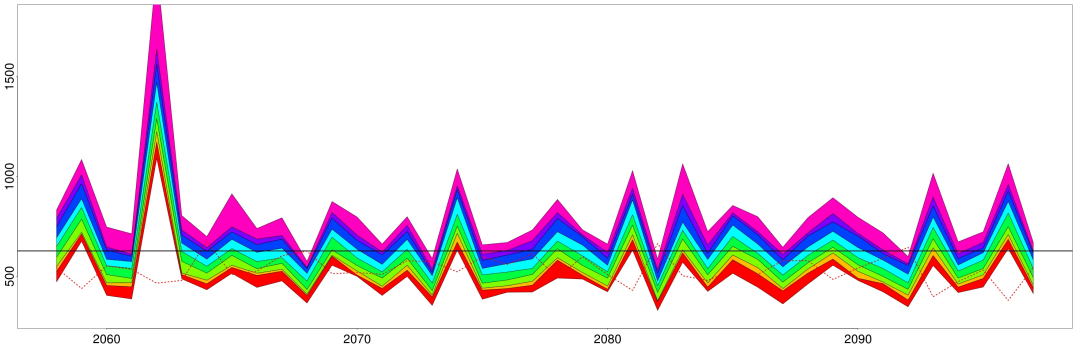
(k) IAP-FGOALS



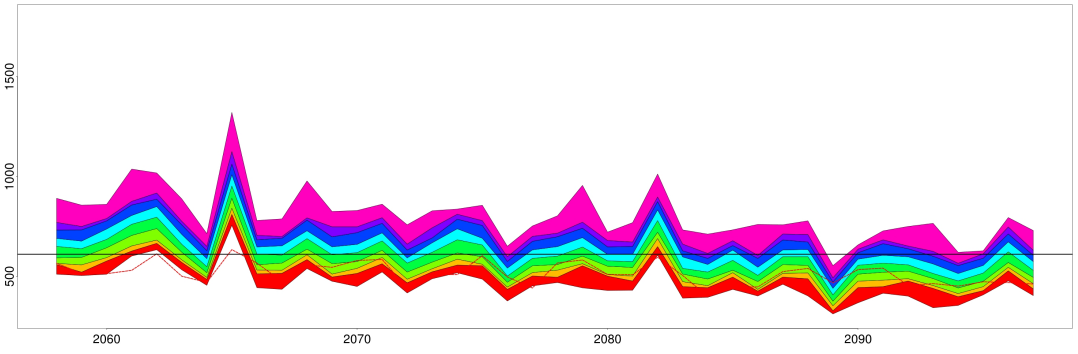
(l) INM-CM3.0



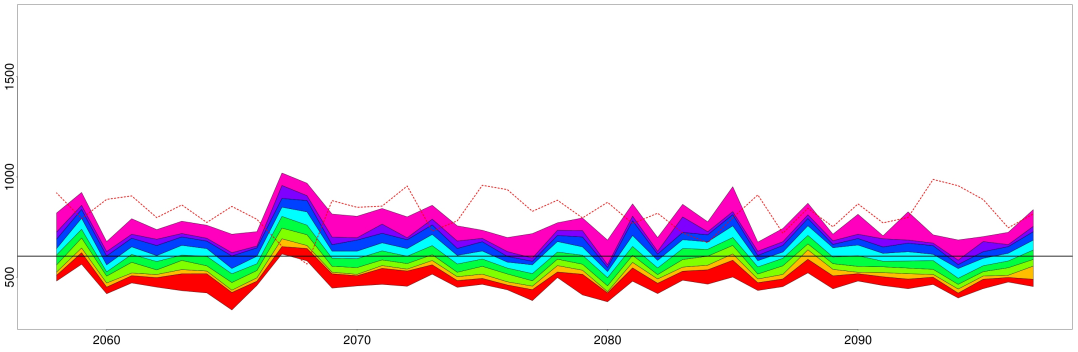
(m) IPSL



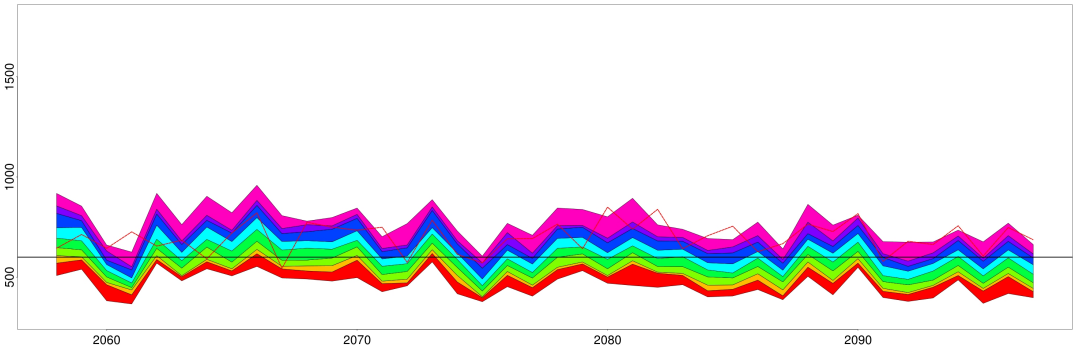
(n) ECHAM5



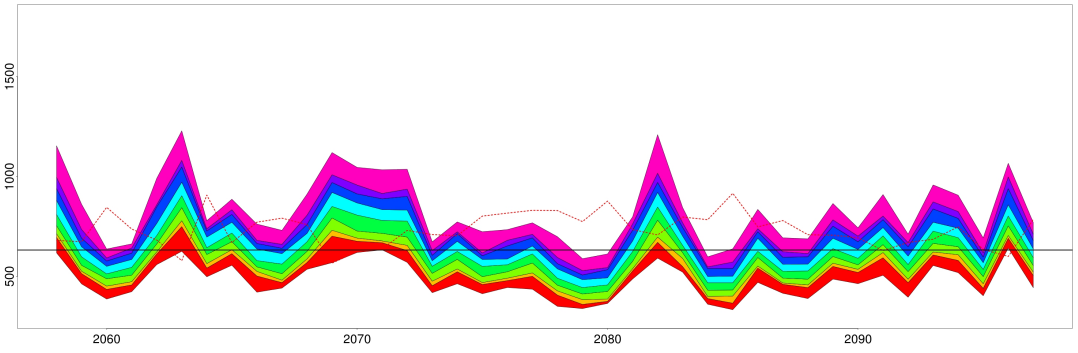
(o) MRI-CGCM



(p) PCM

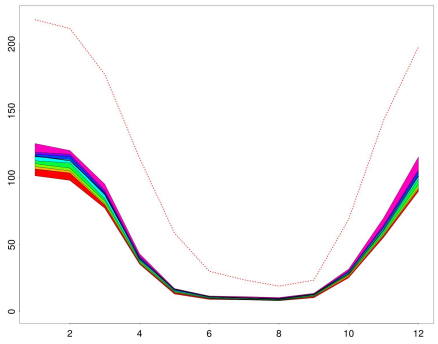


(q) UKMO-HadCM3

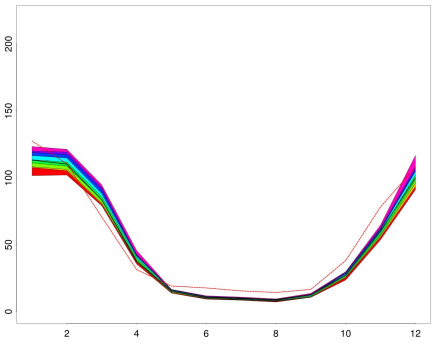


(r) UKMO-HadGEM1

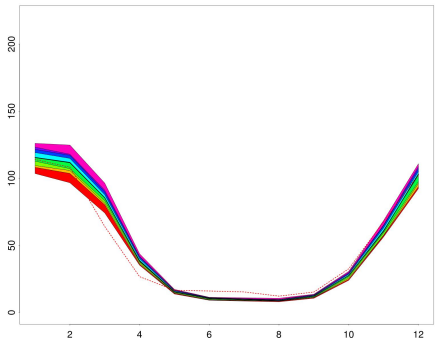
FIGURE 4.4: Rainbow range of simulated monthly climatology and climate model projection (in dashed thin red line) of southern African precipitation averaged over the entire region for the future period and eighteen GCMs. Dashed thick black line indicating overall monthly mean for the corresponding 20^{th} century simulation.



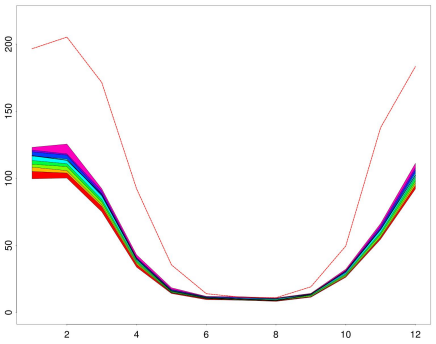
(a) BCCR-BCM2.0



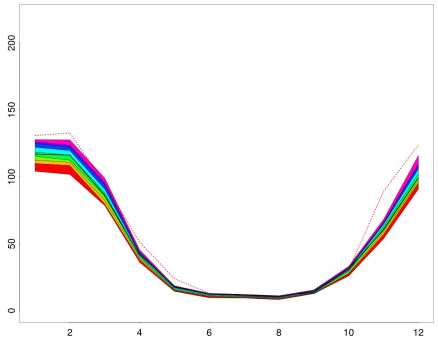
(b) CGCM3.1(T63)



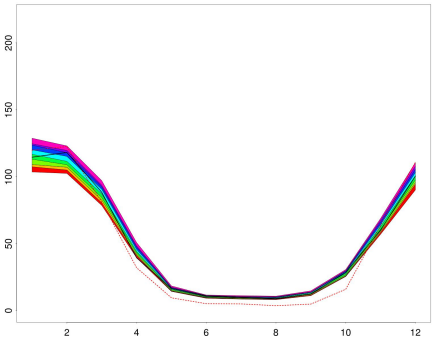
(c) CGCM3.1(T47)



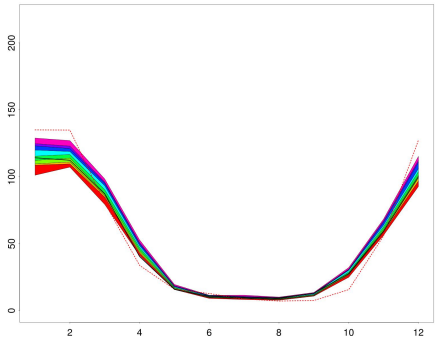
(d) CNRM-CM3



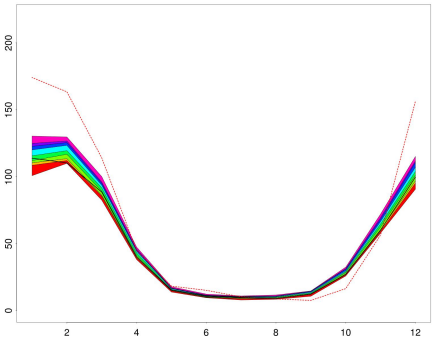
(e) CSIRO-Mk3.0



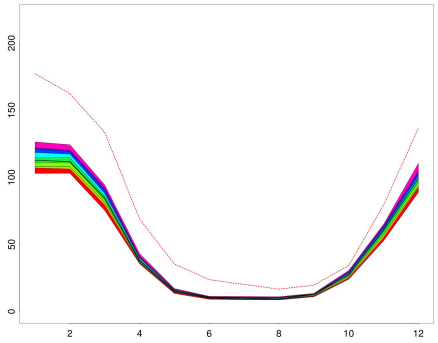
(f) CSIRO-Mk3.5



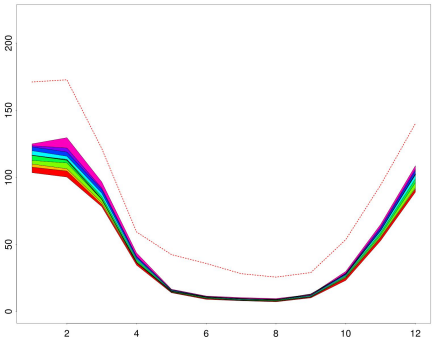
(g) GFDL-CM2.0



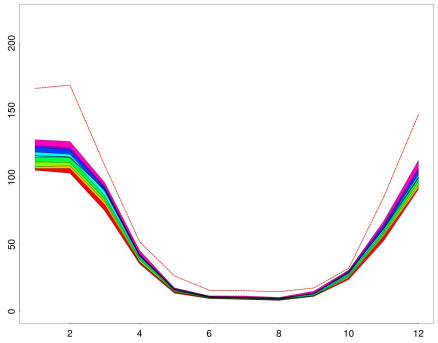
(h) GFDL-CM2.1



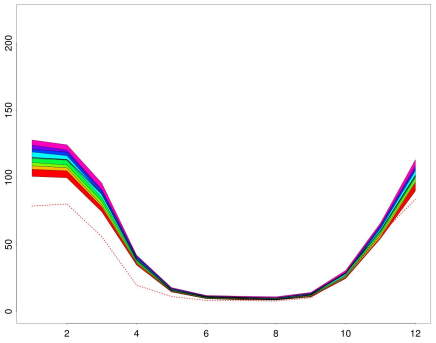
(i) GISS-EH



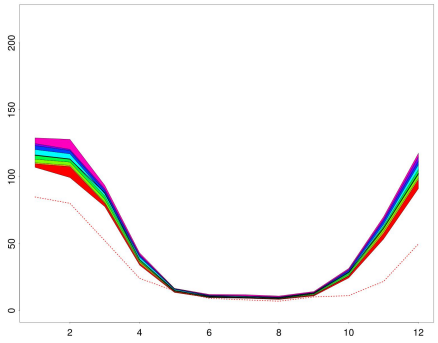
(j) GISS-ER



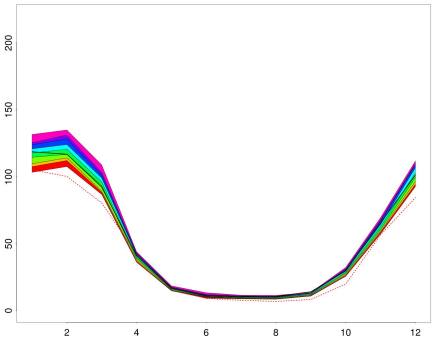
(k) IAP-FGOALS



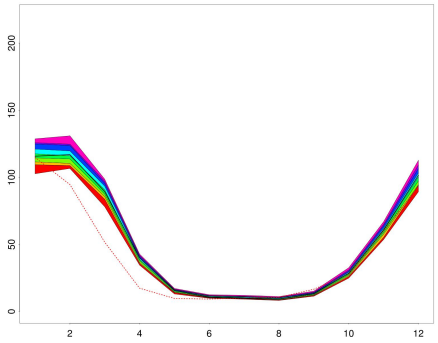
(l) INM-CM3.0



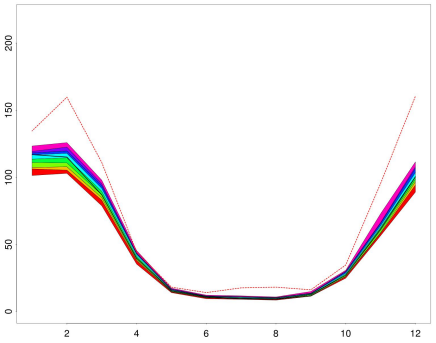
(m) IPSL



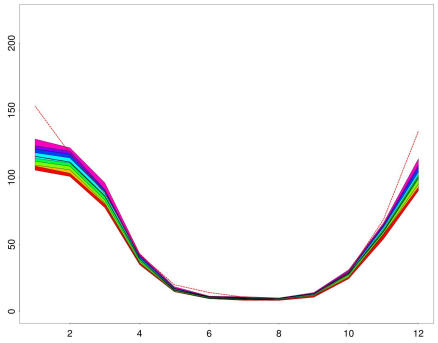
(n) ECHAM5



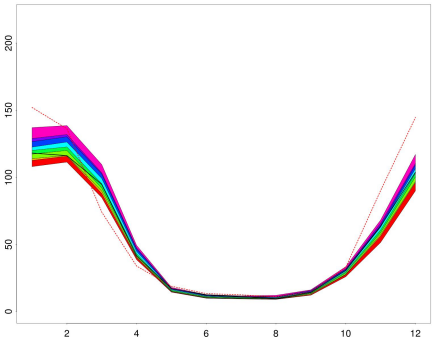
(o) MRI-CGCM



(p) PCM



(q) UKMO-HadCM3



(r) UKMO-HadGEM1

4.4 Conclusion

The application presented here allows the investigation of future projections of southern African rainfall under the probabilistic framework developed in Chapter 3, avoiding the direct use of GCM precipitation fields which are often unreliable.

An initial analysis of the ability of both the introduced statistical framework and the latest generation GCMs in simulating the present day rainfall was carried out at subcontinental level, and at annual and monthly time scales. Compared with the rainfall generated directly by the GCMs, the assessment showed a better agreement between the observed and the statistically simulated rainfall. This supports the use of such synthetic precipitation sequences for subcontinental investigation of precipitation variability and changes under future scenarios, for example.

However, for some of the models the magnitude of the statistically simulated year-to-year variability is lower than that recorded (see for example Figure 4.1, plot i).

The climate models tend to provide projections of precipitation that are systematically biased upwards with respect to the 20^{th} century observations. Moreover, the interannual variability of such projections is rarely consistent with that found in the simulated sequences. This may be related to parametrizations in the climate models not respecting the relationships found in the developed statistical models (e.g. precipitation – relative humidity relationships).

For some GCMs the tendency in over-projecting precipitation, when compared to the simulated sequences, is still visible under a future scenario.

In the considered interval of time during the 21^{st} century, the simulated rainfall sequences show, for some of the climate models, a slightly negative trend. In some other cases a negative trend is visible especially between the 20^{th} century simulations and the final two decades of the 21^{st} century. No clear signal of change is detectable in the monthly seasonal cycle in the simulated sequences between the two centuries.

In this context the GCMs reproduction of the jet stream location proved to be inadequate, resulting in unstable rainfall simulation in the present study. As a result, although an attempt to overcome such problem was made, all the covariates concerning the jet stream were removed from the models (see Section 4.2). Such removal may lead to a slight reduction in decadal variability in the simulated sequences. The effect should

be taken into consideration and solutions when building the jet stream position effect covariate possibly investigated in future research.

In the next chapter a downscaling application of the same statistical framework will be explored for the investigation of precipitation variability at a finer spatial and temporal scale.

Chapter 5

Case Study

5.1 Introduction

The statistical model presented in Chapter 3 was developed to describe large-scale structures in monthly precipitation time series. However, for many impacts studies it is necessary to consider precipitation sequences at finer spatial and temporal scales. As noted previously, the simulation performance of the model was poor at local spatial scales; its monthly time step was a further limitation. Therefore, the application of the methodology to a higher spatial and temporal resolution case study is motivated by the interest to understand the extent to which physical relationships identified between large-scale processes and monthly regional precipitation can be used to study crucial changes within the wet season in a narrower selected area for impact study applications. In this way the large-scale atmospheric and oceanic variables (climate predictors) are used in a statistical downscaling framework for daily precipitation at selected sites, from which growing season metrics are later computed.

The selected narrower area is located in the Greater Sekhukhune district, in the Limpopo province in northeastern South Africa (see Figure 5.1). The study area is prone to severe drought, flooding and significant intraseasonal variability during the wet season (Levey and Jury, 1996; Tennant and Hewitson, 2002; Cook et al., 2004), hence the study of such variability is of more interest than just the seasonal rainfall total (Ziervogel and Calder,

2003). The use of daily data allows the investigation of intraseasonal variation and rainfall characteristics that are relevant in agricultural and other rainfall dependent sectors, where a water shortage of just a few days may result in extensive economic and social damage (e.g. a dry spell during the flowering and the grain-filling stage of maize). The rainfall regime of the region in question is unimodal, mostly from December to February and the crop production generally rainfed. However, relatively few rainfall events may occur during the rainy season and the area has been affected by droughts (2001 – 2004) as well as extreme flooding (2000) in recent years (Reason et al., 2005). Agriculture is still the major source of income and livelihood for most of South Africa's population, and cereals and grains are among the country's most important crops. The largest locally produced field crop and the most important source of carbohydrates in South Africa is maize (*Zea mays* L.) (Smale and Jayne, 2003; Byerlee and Eicher, 1997). As well as extensive monoculture, maize production also dominates the smallholder farming system in the study area (Dimes and du Toit, 2008). It is therefore critical to understand the year-to-year changes in the planting season and rainfall characteristics in order to introduce management decision and mitigation measures in the agricultural sector (e.g. planting of drought-resistant crops or the choice between long and short-season cultivars). Indeed, according to Lauer et al. (1999) one of the strategies that may be easily introduced by farmers is shifting the crop planting dates to adjust to changes in the rainfall regimes.

Following Tadross et al. (2007) and and Raes et al. (2004) a number of indices have been used in order to study the most important aspects influencing the stages of maize growth (see Section 5.5) and their changes under climate change scenarios.

The growing season (GS) is considered as the optimal period for a crop to develop and grow. In a region where the precipitation is markedly seasonal it is extremely important to identify the right moment to plant, avoiding false starts of the rainy season and subsequent crop failures. The maize plant is quite hardy and adaptable to harsh conditions. However, erratic precipitation, if planted too early, or extremely intense rains, if planted too late may lead to insufficient moisture for the germination or seed loss because washed away (Reason et al., 2005). In addition, warmer temperatures and lower levels of precipitation, that may be experienced in an altered climate, could have impacts on the growing and development of the crop, thus detrimental effects on the

yield, increasing food insecurity in the region (Akpalu and Hassan, 2008). These are the reasons that motivated the work presented in this chapter.

In Section 5.3 the GLM framework is used to characterise the relationships between local-scale daily station data and the large-scale components of the climate system identified in the subcontinental study (Chapter 3). After evaluating the simulation performance of the developed models (Section 5.4), indices relevant for the characterization of the growing season were computed for a validation period and sites (Section 5.5.1). Before investigating the projected future changes in such metrics (Section 5.5.3), 20^{th} century simulations of rainfall and growing season indices, conditional on GCM fields, were examined for possible biases due to the chosen GCM (Section 5.5.2). The outputs of this study may be used to inform local farmers about the likely changes in the maize growing season under a selected future climate scenario. Such information can be useful to inform future possible mitigation and adaptation measures to prevent damaging impacts on the crop development.

5.2 Data used

Within the study region, lying between 23.5°S – 25.5°S and 29°E – 31°E and covering an area of roughly 300×300 km^2, daily rainfall records from fifty-one stations have been selected from the NOAA NCDC Daily Global Historical Climatology Network (GHCN) (see Figure 5.2). The data span the period 1^{st} November 1903 to 29^{th} June 2001. However, only two stations have data after 2000 (see proportion of available data in Figure 5.3). Altitude, latitude and longitude information are available for each station. The altitude ranges from sea level to a little above 2000m, with the majority of the stations being located around 1500 – 2000m a.s.l. (see Figure 5.4). A mapped altitude value, computed from the topographical map, as an altitude average for the area surrounding the site has also been generated for each station and appeared useful to reflect the rainfall structure at a site. The topographical map is derived from a 1×1 km^2 gridded product, known as GTOPO30 and available from the U.S. Geological Survey website (`http://eros.usgs.gov/#/Find_Data/Products_and_Data_Available/gtopo30_info`).

As well as obtaining the mapped elevation of each station, the topographic map was used to obtain information on station aspects and topographic variability within the

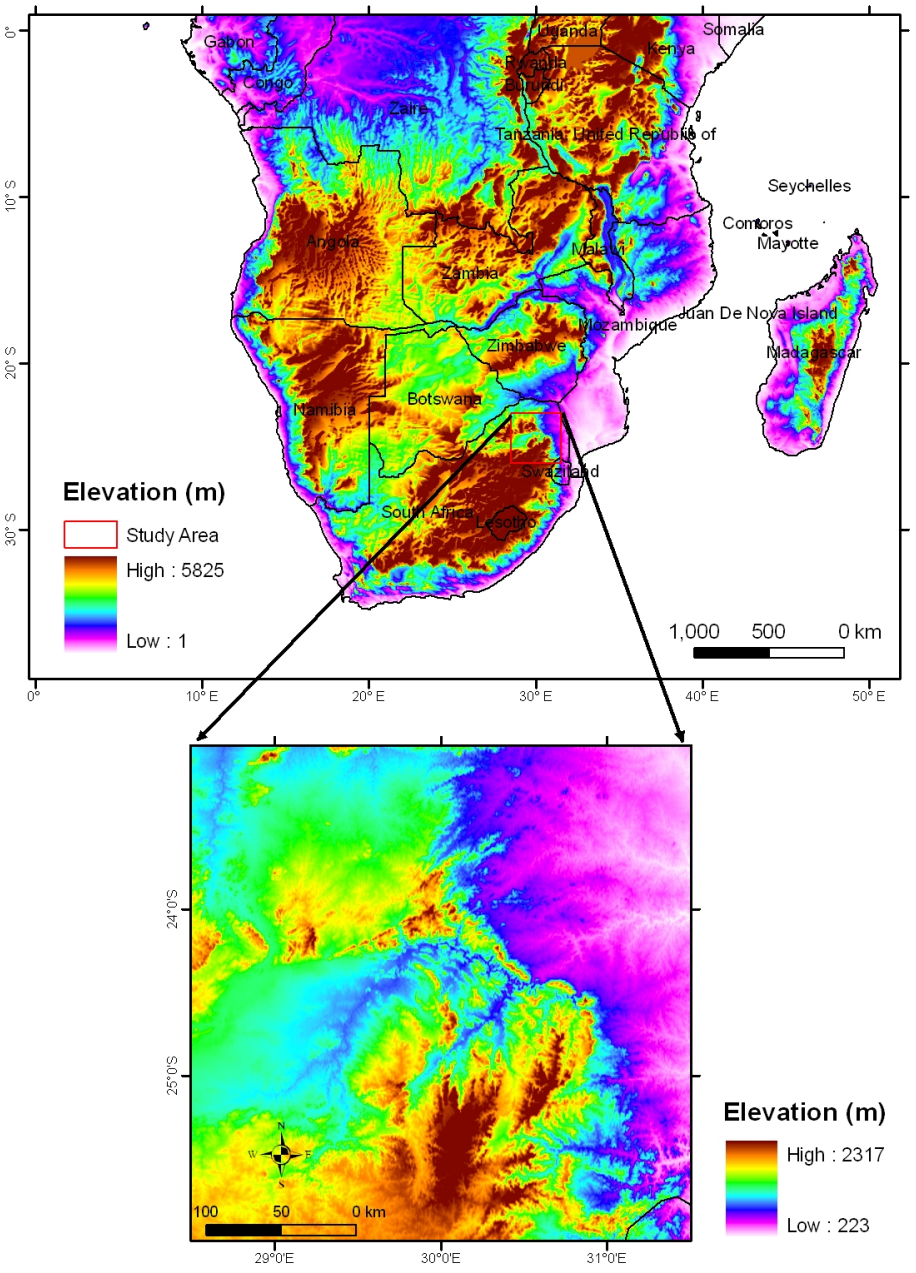


FIGURE 5.1: Location of the Greater Sekhukhune district, Limpopo province, in South Africa (inset map), within southern Africa south of the Equator (big map).

vicinity of each station. This was done by regressing altitude upon both eastings and northings, to produce south-north and west-east slopes information, and computing the altitude standard deviation (topographic variation; see Figure 5.5) for domains in size $3 \times 3\text{km}^2$, $10 \times 10\text{km}^2$ and $30 \times 30\text{km}^2$ centred on each location of interest.

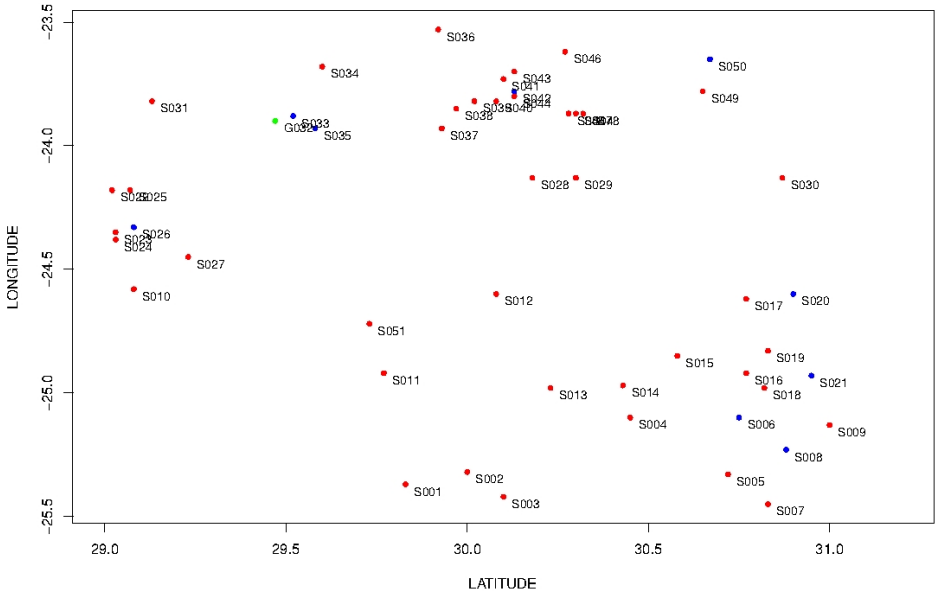


FIGURE 5.2: Locations and codes of the considered weather stations. In blue the stations used in the validation process, in green the discharged stations and in red the remaining ones.

5.2.1 Quality control

Errors due to human mistakes and recording inaccuracies (e.g. record of '0' confused with missing value, inconsistent data series, inhomogeneities as a consequence of instrument deterioration or replacement, variations in the time of observations, and changes in the surrounding environment) often lead to inaccurate or unrealistic precipitation values and characteristics (Beaulieu et al., 2007; Vicente-Serrano et al., 2010). Prior to any detailed analysis therefore, the rainfall data were quality controlled in order to remove common problems potentially occurring with rainfall station data (Berkouwer, 2010).
All the stations have been checked for the recorded precision and proportion of rainfall above 0, 0.5, 1, 2 and 5mm. The result of such quality checks shows large differences in threshold exceeding at neighbouring sites, as well as differences in recording resolution. To ensure a consistent resolution at all sites, the records were all rounded to a resolution of 0.5mm. Moreover, the station coded G032 (green in Figure 5.2) was discarded due to the limited number of observations and the recording of integers only. A threshold, of 1 mm/day, for defining 'wet' days has also been introduced to address inconsistencies

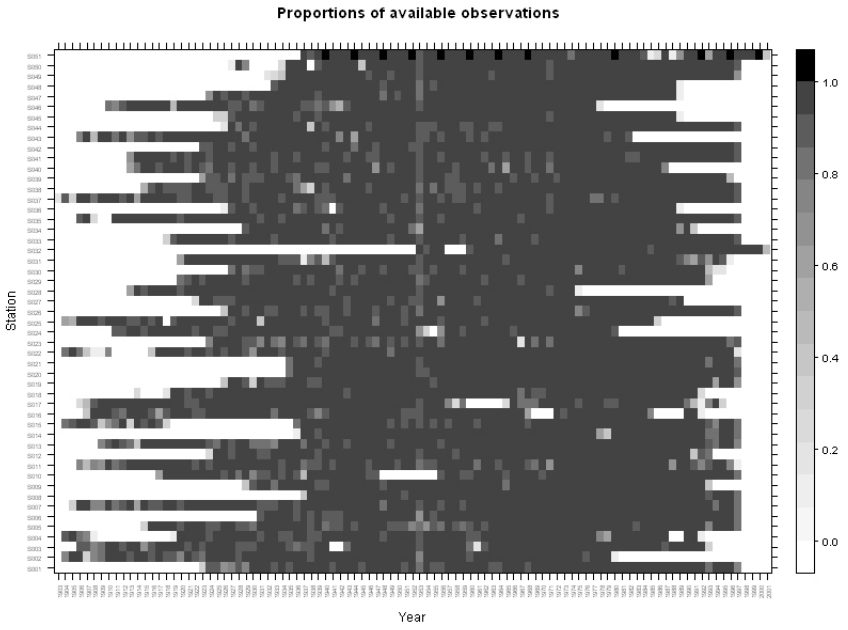


FIGURE 5.3: Data availability for 51 stations in the study region.

among the accuracy of the records. For the source and description of the time-varying data sets used as external climate drivers refer to Section 3.2.

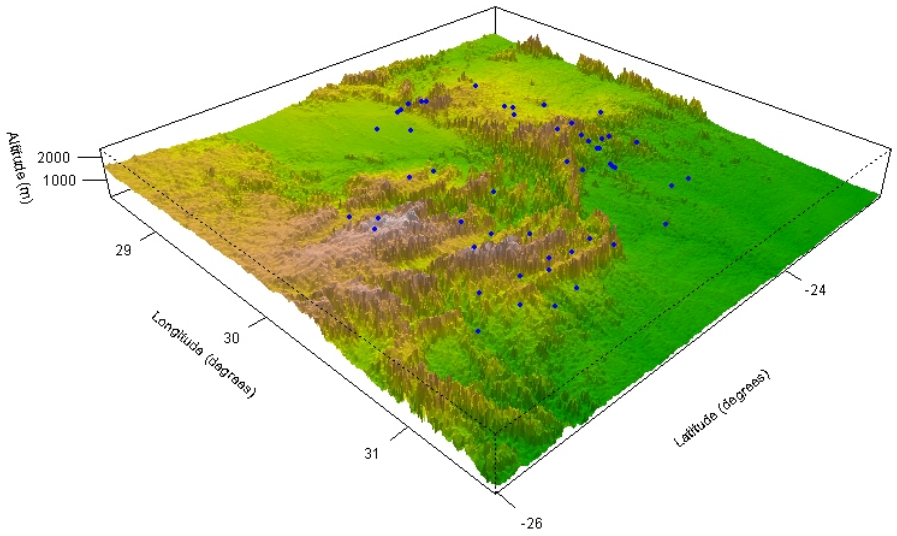


FIGURE 5.4: Topography of the Sekhukhune District. Brown colour represents high altitude and green low. Blue dots represent the locations of the considered 51 weather stations.

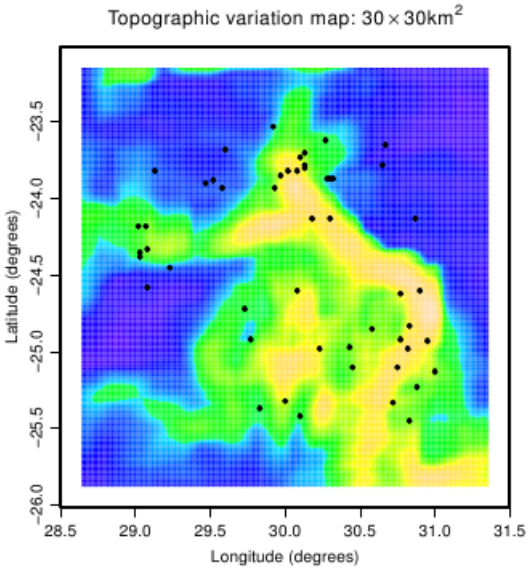


FIGURE 5.5: Topographic variability computed at a spatial scale of around 1000km^2. Blue color represents lower variability and yellow higher.

5.3 Methodology

In Chapter 3, GLMs were used to characterize rainfall variability and to identify large scale factors affecting this variability at monthly and subcontinental scales. The same methodology is now used to investigate daily rainfall in the study area.

The GLIMCLIM (Generalised Linear Model for daily Climate time series) software package (Chandler, 2002) was employed to analyse such variability at finer spatial and temporal scale. Being faster and easier to use when handling large daily data sets, the software was used instead of R (R Development Core Team, 2006). As for the subcontinental study, the first stage was to develop 'baseline' models representing the climatology of the region, for both occurrence and amounts of rainfall. Subsequently, monthly time-varying climate indices were introduced as external covariates. The covariates considered are the same as those used in the subcontinental models, retested here for significance. The reason that led to test such climate factors in this exercise was to investigate which climate drivers, found to explain part of the continental rainfall variability, are significantly linked to local rainfall variability. The use of monthly climate covariates as controls of daily rainfall was justified by the findings reported by Frost et al. (2006). These authors investigated the effect of temporal resolution of atmospheric variables on daily simulated rainfall sequences, comparing the effects of using daily instead of monthly values of the atmospheric variables. They found that the simulated rainfall distributions generated with daily and monthly climate time series were extremely similar. However, the use of daily atmospheric variables lead to more complex models, hence monthly drivers are preferred. Another reason for not considering daily climate model data is that some of the climate models have only 30 days in every month, causing problems with the temporal alignment of data files. Refer to the following Sections (5.3.1, 5.3.2, 5.3.3 and 5.3.4) for a complete description of the models.

From the rainfall data record a subset of 28 complete years (1957 – 1984), for all the 50 available stations, has been used to fit the models. This period was chosen according to the availability of climate covariate time series. For the final validation of the developed models an independent period of 11 complete years (1985 – 1995) has been used for the 9 stations having the most available data in the two final decades of the 20^{th} century (station codes: S006, S008, S020, S021, S026, S033, S035, S042, S050).

5.3.1 Baseline Logistic Regression Model

A logistic regression is used to model the precipitation occurrence assuming 0 to represent a 'dry' day and 1 a 'wet day'. A threshold of 1 mm/day has been introduced to define a wet day as described in Section 5.2. Such threshold was considere adequate in a region where the ETP losses are estimated around 10-15 mm/day (Shukla and Mintz, 1982). If the original variable of interest is Y and the threshold is τ then models are fitted to Y^*, where $Y^* = 0$ if $Y < \tau$, $Y - \tau$ otherwise. After the simulation the threshold is added back to any non-zero values. Therefore the simulated rainfall will not contain values between zero and the threshold (Chandler, 2002). This procedure leads to a slight increase in the amount of rain during wet days. However, the percentage of wet days decreases. Moreover, due to the low threshold the final overestimation of the model should be small enough to be considered negligible.

The baseline model includes factors representing site, daily effects, autocorrelation and spatial dependence. Even though the study area is not particularly large, the complexity of the topography requires the inclusion of several covariates representing the spatial structure. Latitude and longitude have been represented using Legendre polynomials up to the 4^{th} degree and two terms were used to represent the altitude effect upon precipitation: the mapped altitude and the topographic variation, computed over 30×30 km^2 area (Figure 5.5). A west-east slope (over 30×30 km^2) term also proved to be significant, possibly due to the moisture flow coming primarily from the South Indian Ocean. By extracting the longitude and latitude terms from the basic logistic regression model it is possible to show the modelled site effects defined by the Legendre polynomials (Figure 5.6) (Chandler and Wheater, 2002). After adjusting for the topography, the regional structure modelled by the occurrence model shows more wet days in the south-east corner of the study region. The contours represent fitted surfaces as multiplicative contributions to log odds of the 'baseline' level and should not be overinterpreted outside the gauge locations due to edge effects associated with the model representation of regional variation.

The temporal autocorrelation structure has been included using a combination of terms: rainfall values at lags of 1 and 2 days and a 'persistence indicator'. The previous days' occurrence indicators have been transformed using a distance-based weighted average

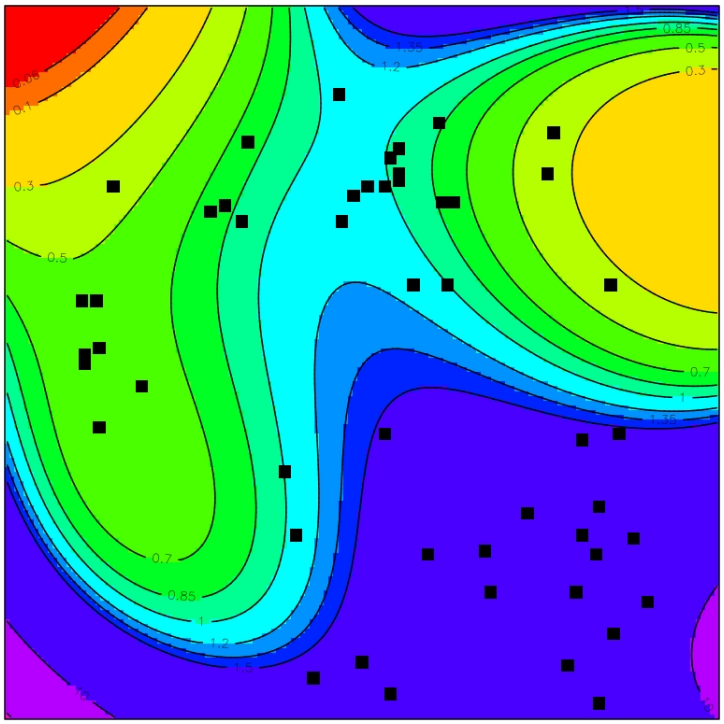


FIGURE 5.6: Site effects modelled by the baseline occurrence model. Contours are regional adjustments to log odds. Red colour indicates low probability of rainfall, purple high probability of rainfall and black squares represent the station locations.

across all sites, where the weights decrease with distance from the current site. The indicator at current site s and day t can be expressed as:

$$\Sigma_r w_{r,s} I\left(Y^{(r)}_{[t-k]} > 0\right)$$

where the weights $w_{r,s}$, associated with site r when predicting for site s, sum to 1 and are proportional to $\exp[-ad_{r,s}]$ (Chandler, 2002). As stated in Chandler (2002), because weather systems tend to affect whole areas rather than only single sites, averaging over all sites can produce better results from a physical perspective.

A 2-day 'persistence indicator' has been added to describe the duration of a weather state for 3 consecutive days, taking value of 1 if the current site was wet on both the previous 2 days, 0 otherwise. Finally, a seasonal structure is represented using Fourier series: sine and cosine for the annual cycle and first harmonic.

As previously stated, due to the software fitting under the assumption of site independence, a spatial dependence structure has to be defined for the model to adjust the

coefficient standard errors and likelihood ratio tests when the locations are not widely separated in space and the sites may be dependent.
In addition to main effects, significant 2-way and 3-way interactions were added to the model. Those interactions (see equation 3.3) suggest that the contribution of the altitude and the autocorrelation structure varies across the area of study.

5.3.2 Final Logistic Regression Model

To identify large-scale drivers of rainfall variability in this case study, it is natural to consider the drivers of subcontinental-scale variation (Section 3.4.2) as potential candidates. Out of the four climate drivers identified from the subcontinental scale analysis, only two proved to be statistically significant at local scale in the occurrence model: relative humidity (RHUM) and southern hemisphere temperature (SHT). A new RHUM index has been created in order to better represent the climate field affecting the study area. The first loading pattern of the principal component analysis has been replaced by the mean relative humidity field above the area, due to the small scale of the domain and relatively coarse resolution of the NCEP dataset used in generating the index. The SHT index is the same as that used previously.
To check for unexplained systematic structure in the final model, Pearson residuals have been computed for subsets of the data. Figure 5.7 shows the monthly and annual Pearson residual means, across all the sites, with a 95% uncertainty band. Under the assumption that the model is correct 95% of the means should lie within the limits. There is no clear structure in the monthly residual means apart from a sequence of negative values during the dry season (May – October), where the model seems to overestimate the probability of precipitation. In the annual plot there is some suggestion of a multi-decadal oscillation in the means, feature that was already visible in the subcontinental scale study (Figure 3.13) and could not be explained using any of the other climate drivers tested (see Table 3.1).
The residual means by site are shown in Figure 5.8. Thick lines correspond to residual means that are significantly different from zero, which, if the model spatial structure is correct, should account for no more than 5% of the total number. That is not case,

although no clear structure is recognisable. For example, the cluster of sites from site S037 to S048 (in the north of the domain; see Figure 5.2) all have residuals that differ significantly from zero but in different directions. Another example is represented by the site S015 (in the south-east of the area) whose residual mean is significantly different from zero and positive but surrounded by negative residual means (sites: S013, S016 and S017). Such inconsistencies were also experienced in a similar case study for a smaller and flatter area in southern England by Yang et al. (2005). Inconsistencies of this type between neighbouring stations must be due either to very local-scale climatological variations or to differences in observer practice. Since the model already contains a very flexible representation of smooth spatial variation (using Legendre polynomials of order up to four), it is not considered worthwhile to try and improve the fit further.

By extracting the climate terms and their interactions it is possible to visualize the expected modelled occurrences of rainfall.

The effect of the SHT index is a linear function of longitude (not shown here) and stronger on the far west of the area, where the overall effect of the SHT (as contribution of 4 standard deviation of the selected terms upon the linear predictor) can be linked to up to 20% decrease in rainfall probabilities. A similar pattern was found in all of southern Africa (see Section 3.4.2, Figure 3.8).

The modelled effect of the RHUM index varies across the region and seasons. The strongest effect is during the rainy season and in the far northeastern corner of the region. In general an overall increase in the index is associated with an increase of up to 5% in the probabilities of rainfall in the summer months. A similar pattern is found in winter. However, because it is the dry season the effect on the precipitation occurrences is almost negligible. The results are comparable to those obtained in Chapter 3 (Figure 3.6). There the seasonal variation was not as strong as here, but similarly an increase in relative humidity was associated with an increase in rainfall (recall that in the subcontinental study an increase in relative humidity corresponded to a reduction in humidity because of the way the index was defined).

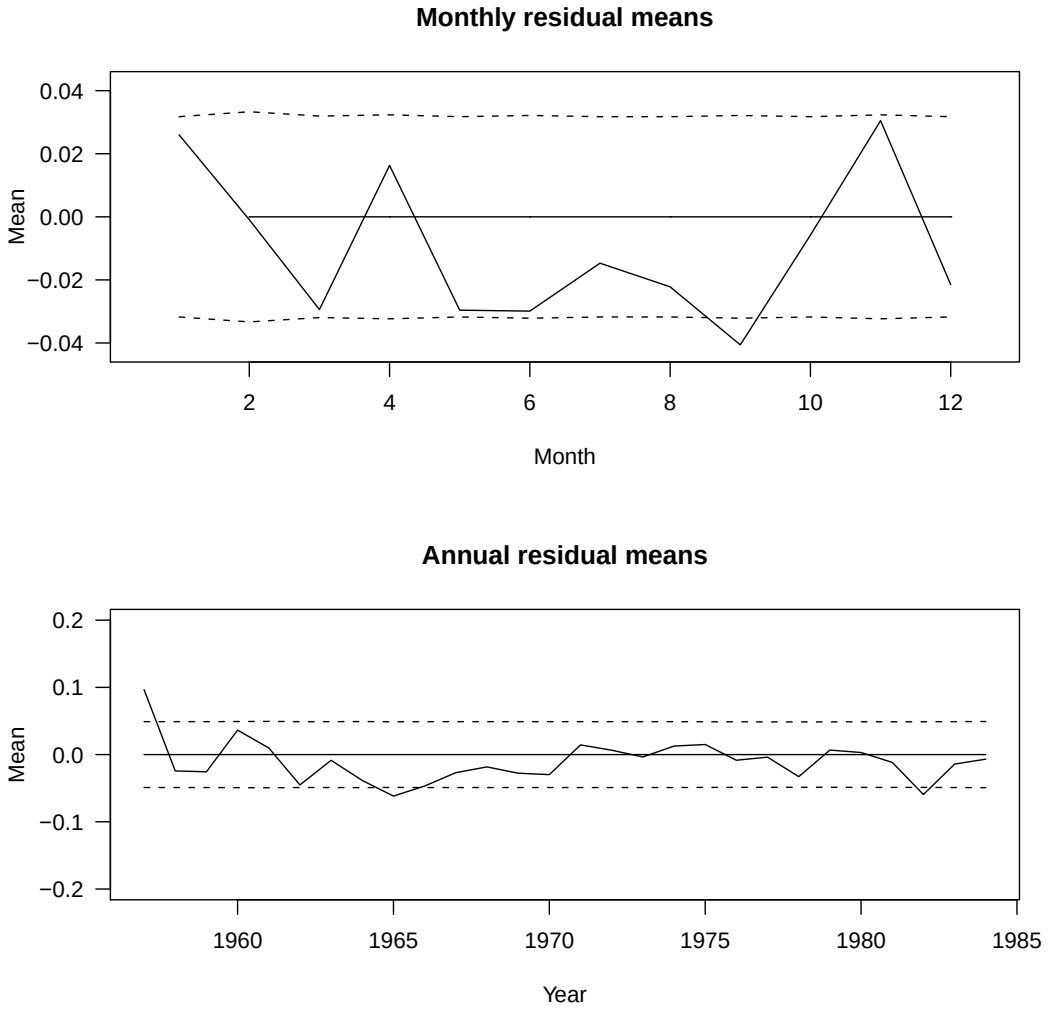


FIGURE 5.7: Occurrence model monthly and annual Pearson residual means across the region. The dashed lines show 95% uncertainty bands under the assumption that the model is correct.

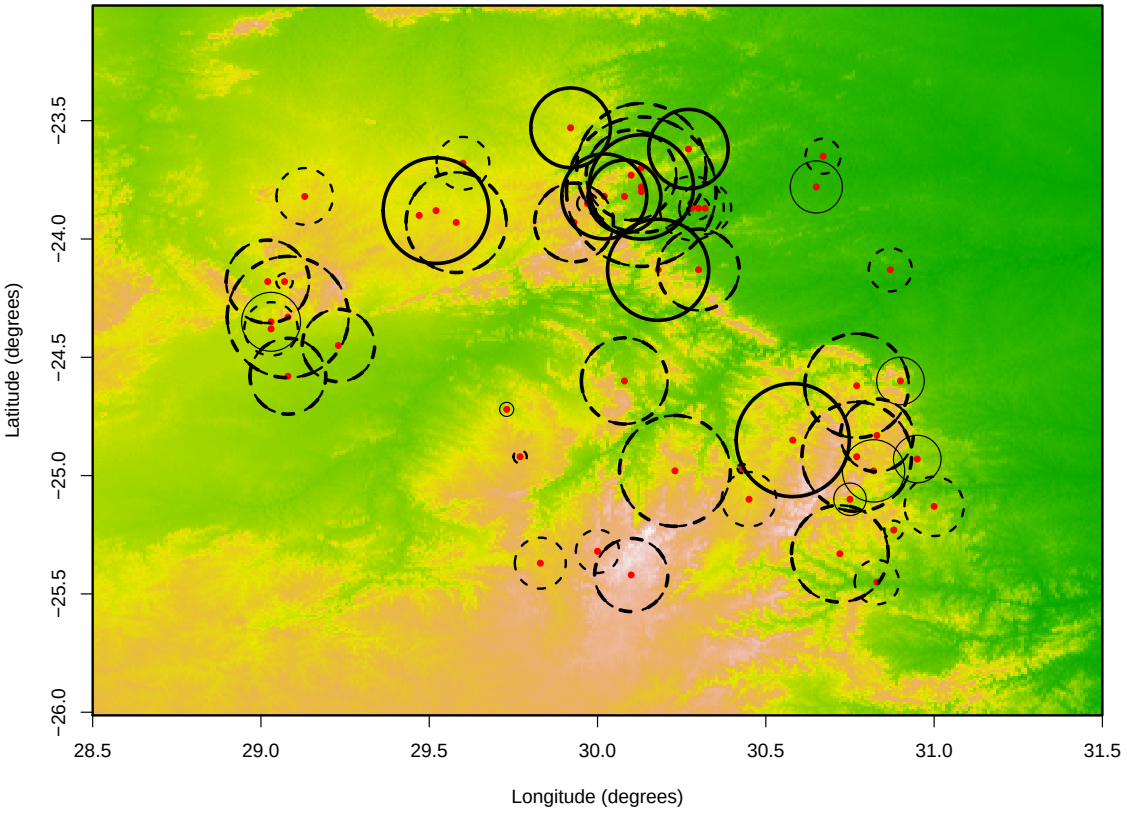


FIGURE 5.8: Occurrence model mean Pearson residuals by site on a topographic map of the region. The red dots represent the station locations and the circles the mean residuals, solid lines for positive and dashed for negative residuals. Thick lines indicate mean residuals that are significantly different from zero at the 5% level.

5.3.2.1 Spatial dependence structure

The dependence in daily rainfall occurrence at a local scale is stronger than at monthly subcontinental scale. An experimental spatial correlation structure, not in the public domain yet, has been implemented in GLIMCLIM in order to subsequently generate rainfall sequences at several sites simultaneously (see Section 5.4). GLIMCLIM generates sets of correlated Gaussian variables $\mathbf{Z} = (Z_1, ..., Z_{S_t})$, where S_t is the number of sites studied on day t. Those pairs of latent Gaussian variables are subsequently thresholded to define correlated binary variables:

$$Y_{st} = \begin{cases} 1 & \text{if } Z_s > \tau_{st} \\ 0 & \text{otherwise} \end{cases}$$

The thresholds $\tau_{st}, ...\tau_{S_t t}$ are chosen to ensure that $P(Y_{st} = 1) = p_{st}$, as required by the logistic regression model, which allows to calculate $E(Y_{st}) = p_{st}$. The correlation between the latent Gaussian variables at each pair of sites is estimated by matching the observed and expected proportions of days for which both sites experience rain. Figure 5.9 shows a scatterplot of the correlation estimated in this way, as a function of inter-site distance. The chosen 'Matérn' correlation structure (Cressie 1991), fitted to the correlations between pair of sites, suggests a rapid decrease in correlations at local scale, leading to highest associations between neighbouring stations, and a slow decay with increasing distance thereafter. The latter may represent a tendency for all sites to be wet or dry throughout the region, suggesting days for which convection tends to be enhanced throughout the region, and others for which it tends to be inhibited.

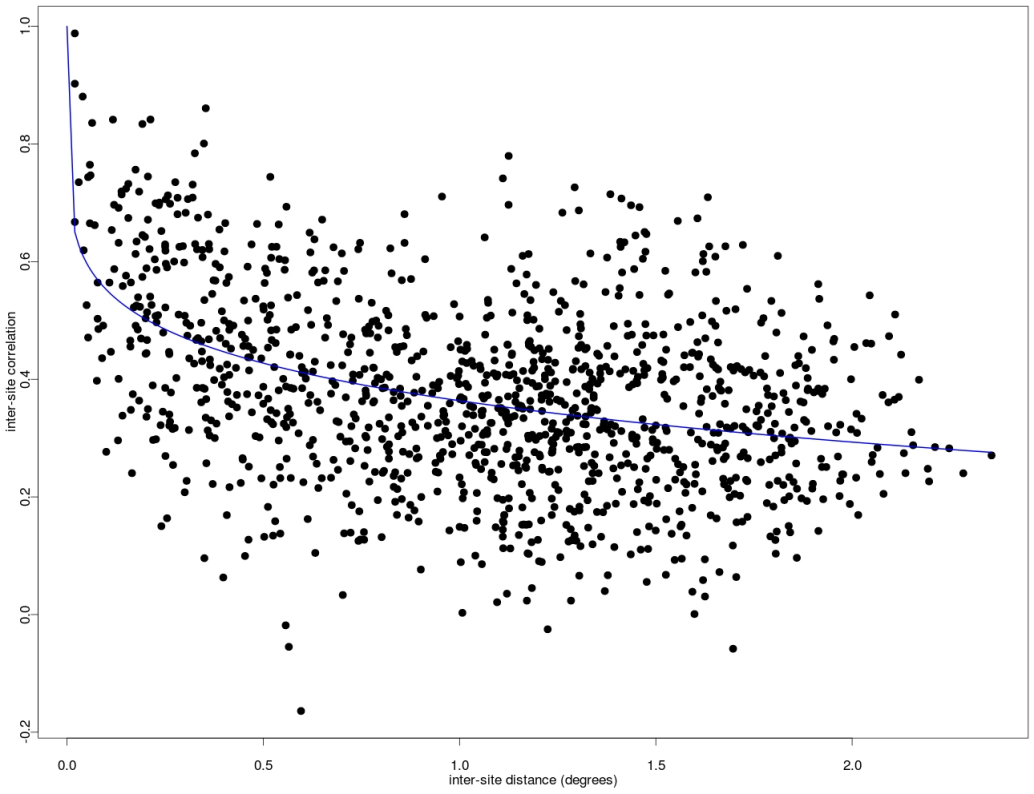


FIGURE 5.9: Matern correlation structure (continuous blue line) for the occurrence model fitted to pairs of inter-site correlations according to their inter-site distances.

5.3.3 Baseline Amounts Model

The steps used in the development of the amounts model follow the same procedure used for the occurrence model and that used in Section 3.4.1.2. A gamma distribution is fitted to the same dataset on 'wet' days only. The same 1 mm/day threshold has been applied.

As for the occurrence model, a large number of terms is employed to represent the complex geography of the area: four degrees of Legendre polynomials, a west-east slope (over $30 \times 30\text{km}^2$) (consistently with the Indian Ocean being the primary source of moisture for the region; see Section 1.2.1) and the same two terms representing the altitude effect, the mapped altitude and the standard deviation of the altitude (over $30 \times 30\text{km}^2$). Similarly to the occurrence model (Figure 5.6), the expected rainfall amounts as a function of the latitude and longitude terms only, are higher in the southeast corner (see Figure 5.10). In addition a northeast-southwest gradient in expected precipitation amounts is also visible.

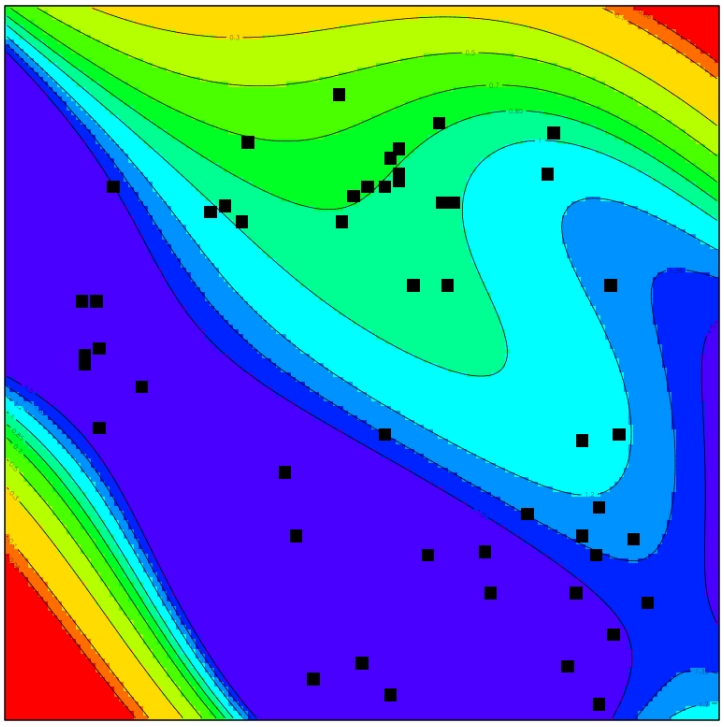


FIGURE 5.10: Site effects modelled by the amounts model. Contours are multiplicative adjustments to log of expected rainfall amounts. Red colour indicates low mean rainfall, purple high mean rainfall and black squares represent the station locations.

The autocorrelation structure is expressed using 1 and 2 previous days rainfall amounts transformed using the same distance-based weighted average across all sites used in the occurrence model (see Section 5.3.1), but for the gamma model (see Section 3.3.6). Similarly, a 2-day persistence indicator and four sine and cosine terms representing the daily seasonal cycle were added to the model. In addition, the model contains a smooth month adjustment for August due to high Pearson residual means previously found under a sine and cosine-wave model only. The adjustment allows to smoothly model a departure from the overall seasonal cycle avoiding unnatural discontinuities, reaching the maximum value of 1 in the middle of the month and smoothly decaying to 0 at the end of the month (Chandler and Wheater, 2002).

The spatial dependence structure is modelled via a residual correlation model similar to the one described in Section 3.3.7. Empirical correlations between pairs of Anscombe residuals (see equation 3.4), approximately normally distributed, are computed and used to generate a multivariate normal random vector of the quantities. The residual transformation is then inverted to obtain rainfall amounts in the simulation process (Yang et al., 2005).

Geographical interactions include latitude and longitude with altitude terms and west-east slope.

5.3.4 Final Amounts Model

With respect to the subcontinental study, time-varying climate indices such as relative humidity, El Niño, SHT and SWIO, and their interactions, retained their significance in explaining part of the variability in rainfall amounts also at a narrower spatial scale and daily time step. After finalizing the model, Pearson residuals were inspected to check for unexplained structure.

The top plot of Figure 5.11 represents the monthly Pearson residual means, where no sign of systematic structure is visible. The increased width of the uncertainty bands from May to September is due to the lower number of observed 'wet' days occurring during the dry season. Similarly annual Pearson residual means present no systematic structure (see bottom plot in Figure 5.11) in contrast with the multi-decadal oscillation found in the occurrence model (see Figure 5.8, bottom plot). Interestingly the lowest

negative residual, which occurs in 1986 (Figure 5.11, bottom plot), is also evident in the annual residuals of the occurrence model (Figure 5.8, bottom plot).

Even though 22 single and 2-way interaction terms are included in the model to represent the spatial variation, negative and positive residual means at neighbouring sites are still visible in Figure 5.12, where the Pearson residual means are plotted by site. However, the spatial inconsistencies are less dramatic than for the occurrence model. As before, the complexity of the geography can not be resolved further without risking overfitting the model.

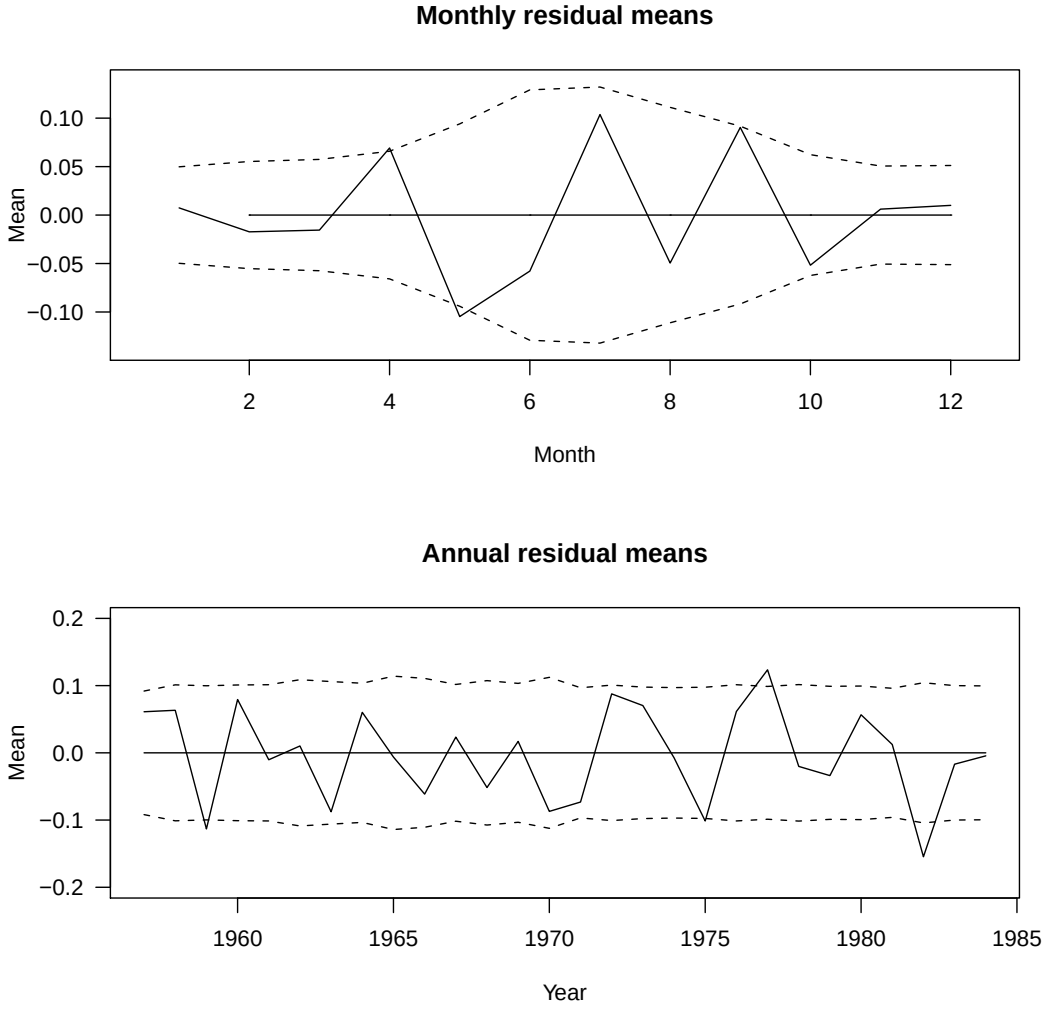


FIGURE 5.11: Amounts model monthly and annual Pearson residual means across the region. The dashed lines show 95% confidence intervals under the assumption that the model is correct.

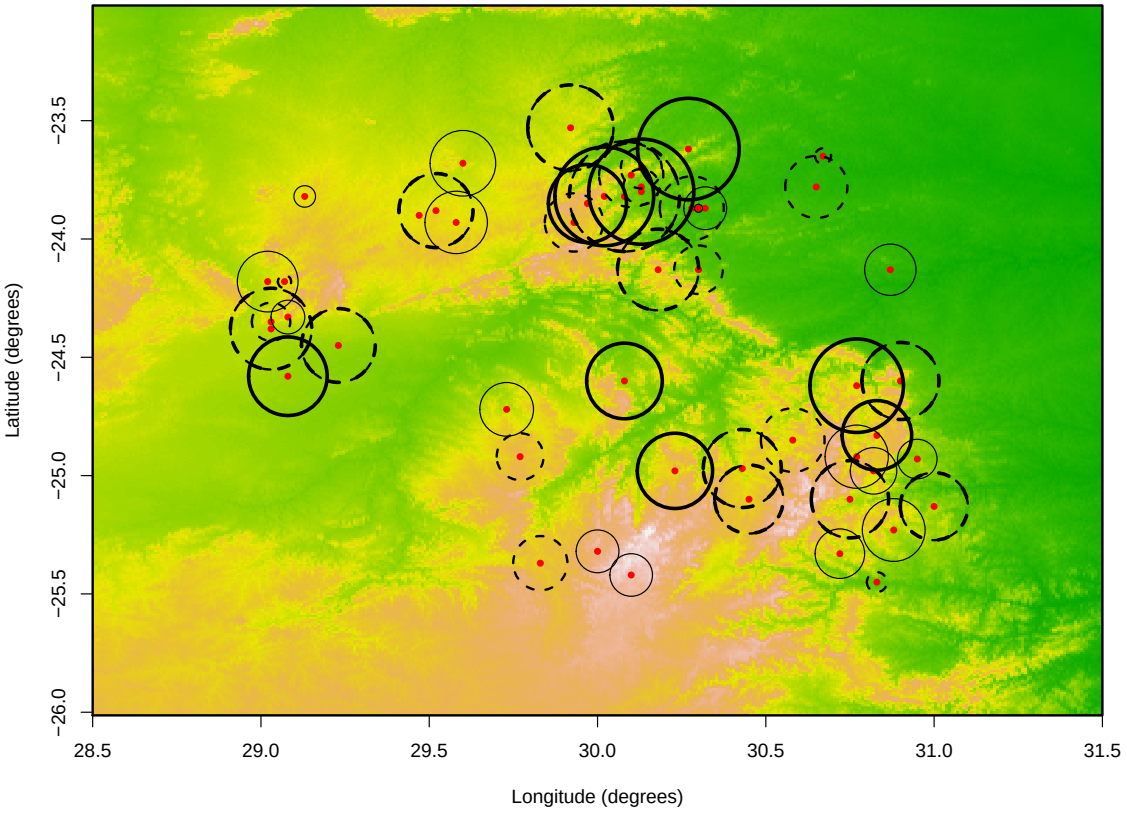


FIGURE 5.12: Amounts model mean Pearson residuals by site on a topographic map of the region. The red dots represent the station locations and the circles the residuals, solid lines for positive and dashed for negative residuals. Thick lines indicate residuals that are significantly different from zero at 5%.

The modelled effects of the selected climate covariates upon rainfall amounts are investigated by plotting the contribution of 4 standard deviations of the selected terms to the linear predictor (not shown here). The effect of the relative humidity index, and its interactions with latitude and longitude, is characterised by a distinct northeast-southwest gradient, with the strongest positive effect on precipitation amounts on the top eastern corner. The modelled response is similar to the effect upon rainfall occurrence, but without seasonal variation. Again this agrees with the results from the subcontinental study (Figure 3.9).

The south-west Indian Ocean (SWIO) index effect involves significant interactions with the seasonal cycle. In order to display their effect upon the seasonal cycle of precipitation, Figure 5.13 shows the modelled seasonal cycle both when the index in set to its average value and when it is 2 standard deviations above average. An enhanced SWIO index (increased north-south gradient in Indian Ocean SSTs) causes a strengthening and a slight delay in the average seasonal cycle, in agreement with the subcontinental scale results from Section 3.13. Both of the two remaining climate drivers incorporated

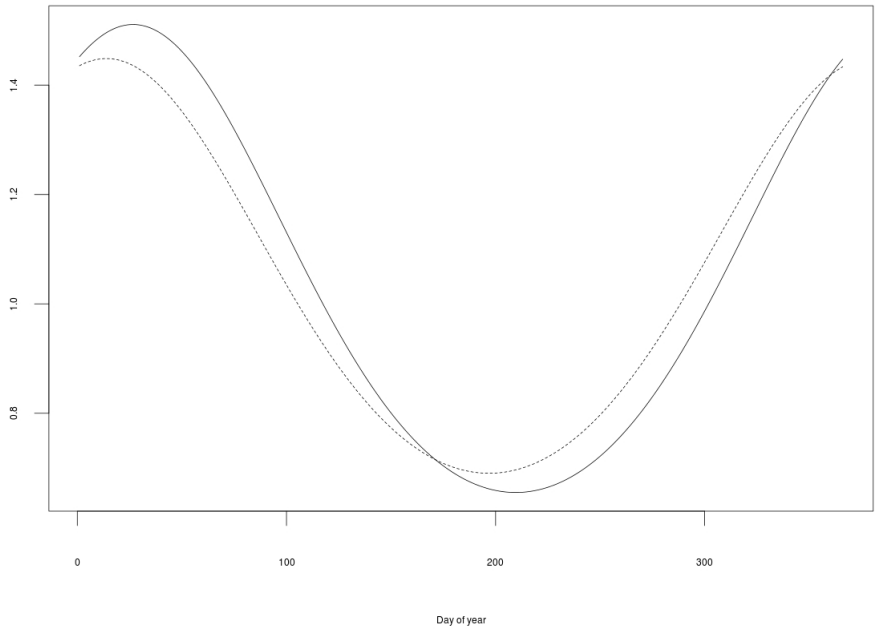


FIGURE 5.13: Effects of the SWIO index and interactions as modelled by the Amounts model. Lines are multiplicative adjustment to an overall mean level. Solid line indicates a 2 standard deviation enhanced seasonal cycle. Dashed line represent an average seasonal cycle.

into the amounts model, El Niño Index and the SHT Index, interact with the relative humidity index. Showing the effect of such indices upon rainfall amounts is complex as the terms may be correlated. However, it is possible to describe the effect of such interactions, assuming that the teleconnections modulate the relative humidity – precipitation relationships, by considering that the corresponding regression coefficients can be written (see definition in Section 3.3.1) as:

$$\beta_{\mathrm{RHUM}} = \beta_0 + (\gamma_1 \times \mathrm{NINO3.4}) + (\gamma_2 \times \mathrm{SHT})$$

where β_0 is the main effect coefficient (0.0399) and ($\gamma_1 \times$NINO3.4) and ($\gamma_2 \times$SHT) are the interaction coefficients with the climate indices, respectively 0.0424 for the interaction between El Niño and relative humidity and -0.0260 for the interaction between SHT and relative humidity.

5.4 Daily rainfall simulation

After checking the model performances, in terms of residuals analysis and modelled effects of some of the covariates upon precipitation, simulation abilities have been tested. The occurrence and amounts fitted models have been used in combination to produce synthetic rainfall sequences for an independent period of time and validation stations. The period used for validation, independent from the fitting procedure, runs from 1st January 1985 to 31st December 1996 allowing 11 growing seasons to be investigated. Only the stations with the most complete record in the last two decades of the 20th century have been used for validation (S006, S008, S020, S021, S026, S033, S035, S042 and S050; see blue stations in Figure 5.2). Nonetheless, the selected records contain a few missing values (less than 2%). To account for this, the distribution of the missing data, conditional upon the available observations, has also been simulated. The imputation procedure, performed by GLIMCLIM, consists of using the fitted models in such a way that the missing data are simulated conditional on all available data so as to produce uncertainty envelopes for the historical rainfall statistics. Although none of the validation sites is located at the centre of the domain, they are situated both in areas of high (southeast corner) and low (northwest corner) topographic variability, and hence represent different geographic conditions within the study area.

Figure 5.14 shows the distribution of observed and simulated seasonal rainfall means for a 9-gauge average daily series. The black band shows the results for the 5%-95% percentile of 100 sets of imputation of missing data. The maxima and minima were omitted to reduce the sensitivity of the validation exercise to simulated outliers. The shaded bands represent the range, along with the 0-5th, 5th-10th, 10th-25th, 25th-50th, 50th-75th, 75th-90th, 90th-95th and 95th-100th percentiles, from 100 simulations. During the winter months and in some years (top plot) the black bands are slightly thicker due to the higher proportion of missing data with respect to the summer months (bottom plot). In the winter months the simulated distribution of annual seasonal precipitation encompasses the observational station data, except for one year, 1988. In that year a strong El Niño event was recorded: its effect can be seen clearly in the simulated rainfall distributions. However, it did not appear to be associated with strong negative precipitation anomalies in the considered stations.

Moreover, the moderate decrease recorded in winter precipitation and visible in the latter half of the considered period, is not matched by a similar feature in the simulated series. Although 11 years is not enough to draw conclusions about interdecadal variability, this is in agreement with the results of the regional scale study (see Section 3.5), where decades of under and over simulation were attributed to the possible lack in the statistical model of some interdecadal variability signal. Because these results concern the dry season however, and the difference between simulated mean and observed precipitation values is only about 20-30 mm/year, the outcome is hardly relevant for the aim of this research.

As far as the wet season is concerned, the bottom plot in Figure 5.14 (the year corresponds to the Jan-Feb months) shows that the range of the simulations cover the seasonal rainfall variability of the observations. Again only one extreme season is slightly outside the simulated range. However, like the observations, the simulated distributions are still shifted downwards compared to the neighbouring years, suggesting the presence of a rainfall reduction signal in the climate drivers included into the statistical framework.

The plots in Figure 5.15 show the observed values and simulated distributions of various summary statistics for seven of the nine selected gauges. The results for the remaining two gauges are more difficult to interpret due to a higher proportion of missing data at these gauges, which leads to wide imputations envelopes. In general there is broad agreement among simulated and observed statistics. The observed monthly means (top left plot) fall within the simulated range. Nonetheless, the model does not reproduce the strange pattern observed towards the end of rainy season (Feb-Mar), a problem also visible in other statistics. In particular the highest discrepancies happen in February, for which the simulated values are generally too low and unable to reproduce the sudden rise seen in the observations. However, that does not seem to be the case in two of the considered stations (Figure 5.15, plots a and b), and was not evident in the monthly residual means checked during the modelling process. This suggests once more that there may be inconsistencies within the region that can not be accounted for by the model.

Other discrepancies between the observed and simulated monthly statistics are visible in some of the plots, however, the overall structure of the observation is reflected in the simulated distributions.

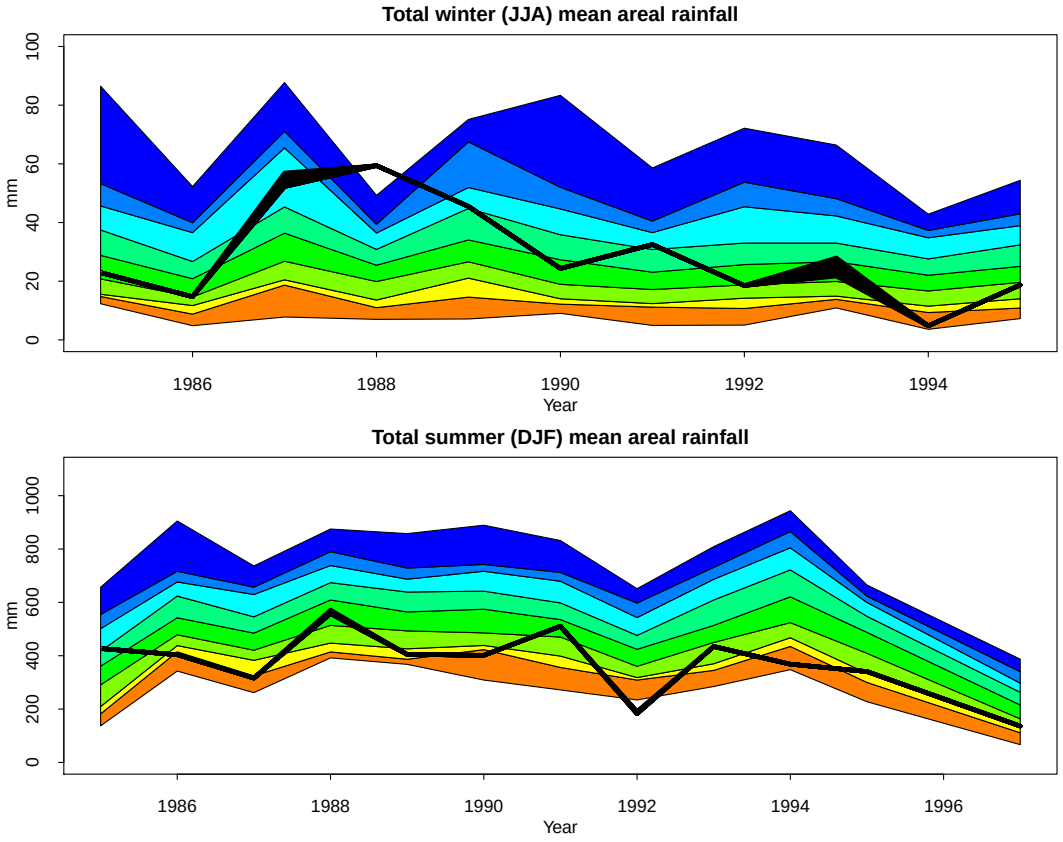
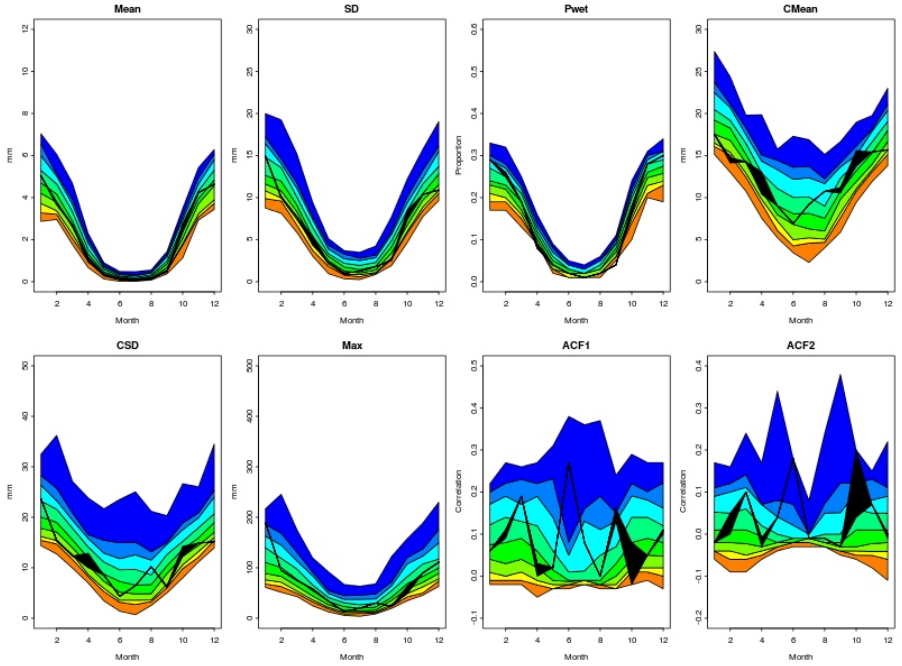
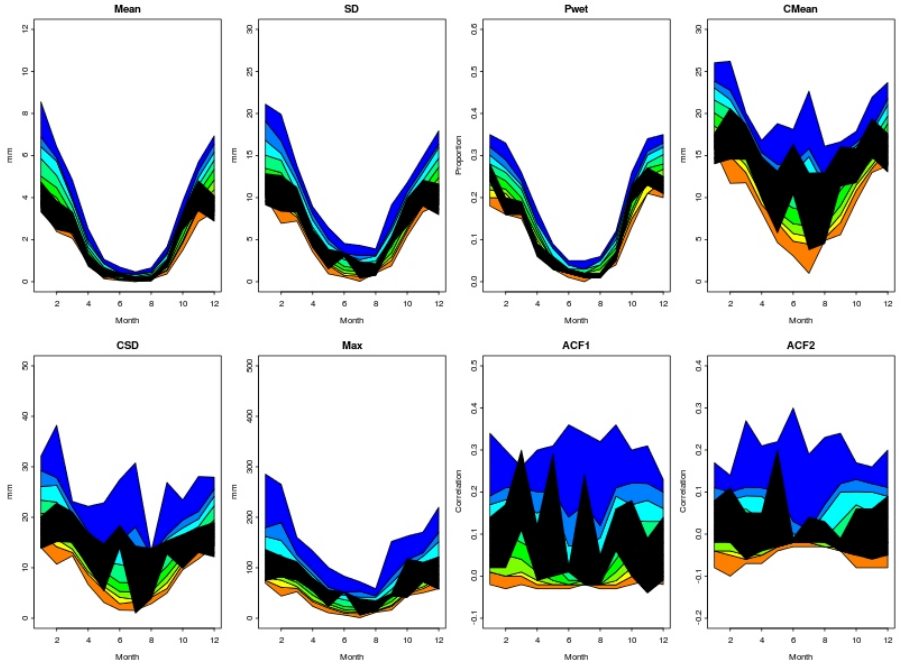


FIGURE 5.14: Observed and simulated seasonal precipitation means for winter (top) and summer (bottom) months averaged over 9 selected sites. The coloured bands represent the range of the distribution and the 0-5th, 5th-10th, 10th-25th, 25th-50th, 50th-75th, 75th-90th, 90th-95th and 95th-100th percentiles of 100 simulations. Thick black lines represent the 5th–95th percentile envelope of 100 imputations.

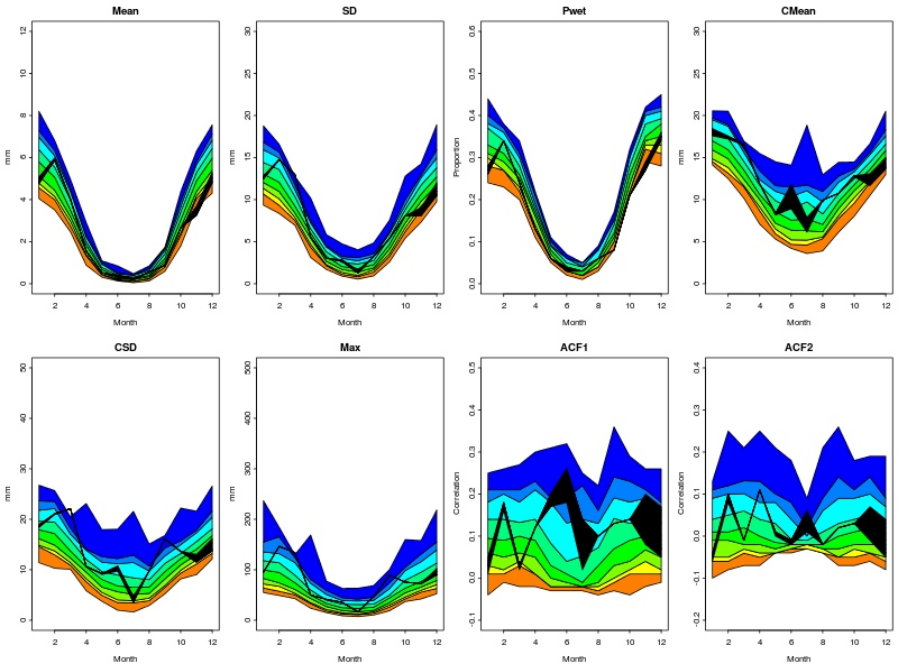
FIGURE 5.15: Observed and simulated summary statistics for 7 of the stations used in the validation process. The shaded bands represent the quantiles of the simulated distribution and the thick black line the 5^{th}–95^{th} percentile of the envelope from 100 imputations of missing data. In the top row are the monthly mean, standard deviation, proportion of wet days and conditional mean (i.e. mean on wet days only). In the bottom row the conditional standard deviation, maximum and autocorrelation at lags one and two. See Figure 5.2 for the locations of the stations.



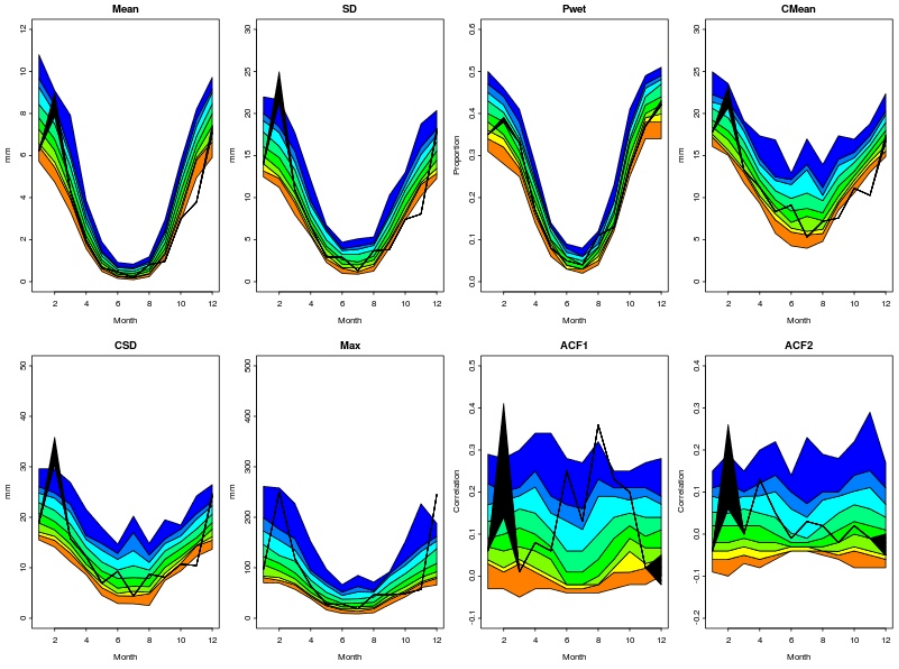
(a) Station Code: S006



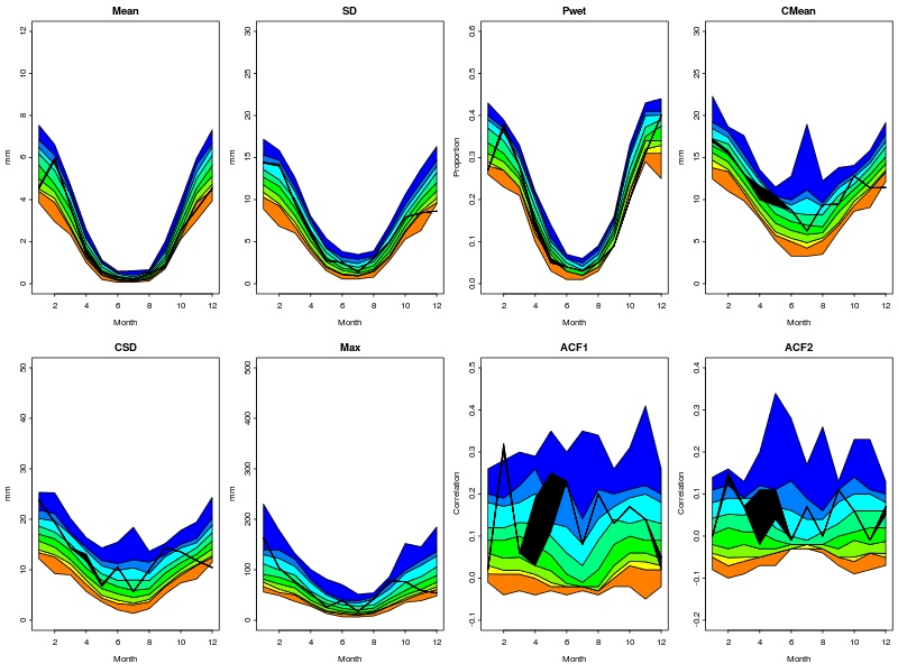
(b) Station Code: S020



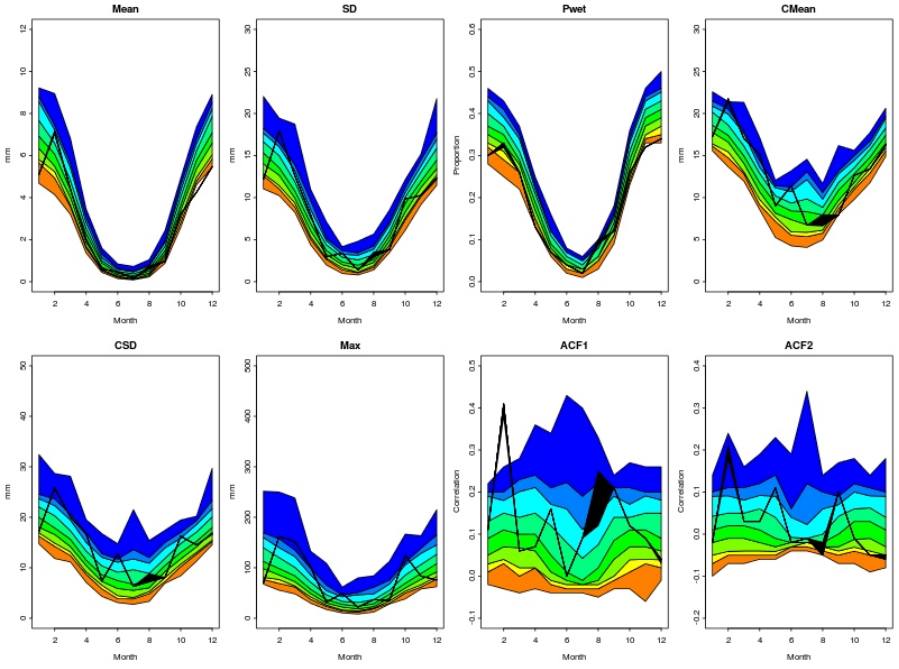
(c) Station Code: S026



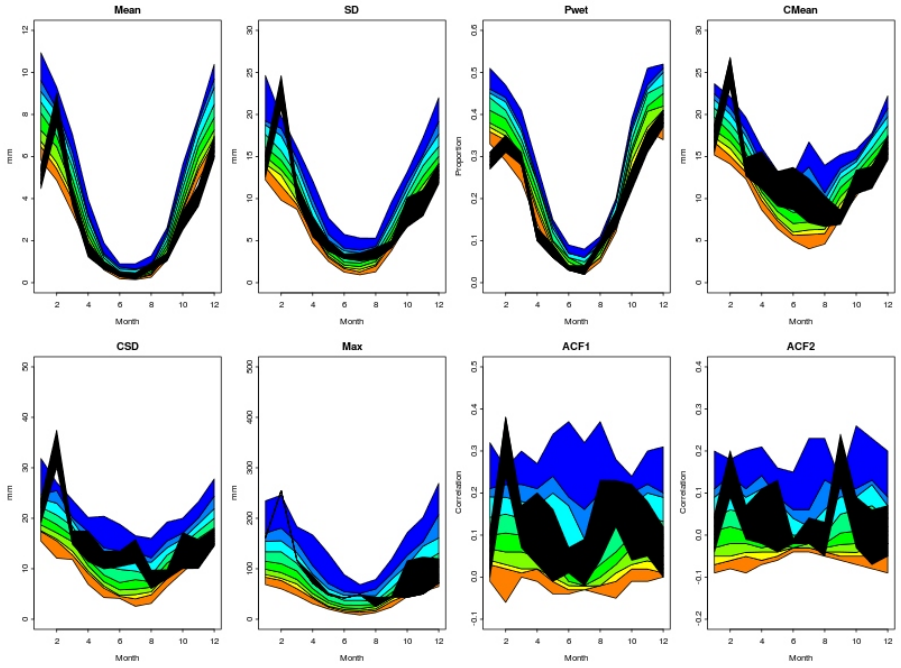
(d) Station Code: S033



(e) Station Code: S035



(f) Station Code: S042



(g) Station Code: S050

Finally, to verify the suitability of the spatial correlation structure in the occurrence model the simulated distributions of the number of wet sites have been checked (Figure 5.16). Since neighbouring sites tend to experience the same weather state, an adequate spatial correlation structure should adequately reproduce the observed frequency of wet sites. The plots show, for January and July, the observed frequencies of wet sites in green and 5% and 95% percentiles of simulated distribution in red and blue respectively. There is generally broad agreement between the observed and simulated distributions of number of wet sites. However, during both the wet and dry season, the model does not represent enough days when all sites experience no rain. Further investigation on the used spatial dependence structure may be required.

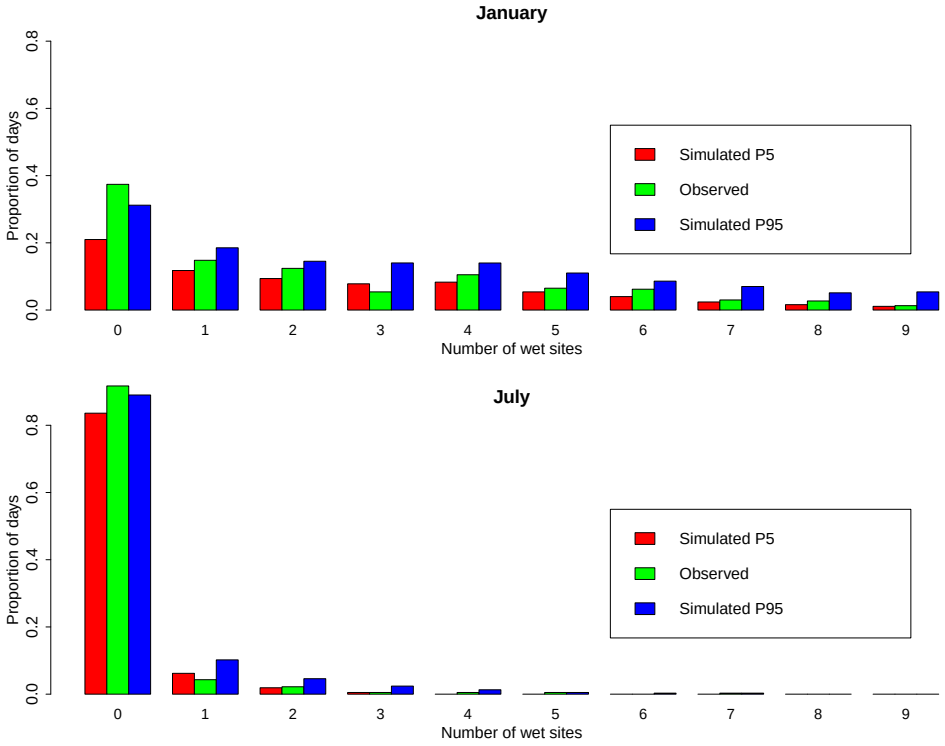


FIGURE 5.16: Observed and simulated distribution (5^{th}–95^{th} percentile) of number of wet sites (out of 9) for January (top) and July (bottom).

5.5 Interannual variability and rainfall characteristics during the growing season

After examining the performance of the models in reproducing generic features of rainfall series, the present section will focus on metrics that are specifically relevant to the impact applications that originally motivated the study.

A list of indices, previously suggested and applied in the African continent, has been selected in order to investigate some aspect of intraseasonal rainfall variability having an impact on maize cropping (Table 5.1). The start date of the growing season is critically important for the farmers as they need to determine when is the best date for planting the crop. Previous studies have defined the onset using a variety of indices as the interpretation on how much rain is considered enough for a crop to be planted safely is not unique, rather crop and region-dependent (Raes et al., 2004; Tadross et al., 2007; Hachigonta et al., 2008; Kniveton et al., 2009). As reported by Kniveton et al. (2009), Stern et al. (1982) mentioned a general definition involving the occurrence of at least a certain amount of rain in a certain number of days after a certain date. A further criterion is introduced in most indices to avoid incurring in a false start: it is considered a false start when a dry spell of certain number of days occurs after the considered onset. However, following the definitions given in Raes et al. (2004) and Tadross et al. (2007), the onset of the growing season is identified in the present work by the first four days in which at least 40 mm of rain falls. This criterion was initially introduced in Raes et al. (2004) based on farmers' practices in order to reduce the risk of planting failure. No criterion is introduced to avoid false starts because, although such criteria are useful for retrospective analysis, they are irrelevant to guide farmers in their choice (Ati et al., 2002).

The other indices in Table 5.1 include the length of the growing season: this is an indication of the optimal crop growth period. The characteristics of dry breaks (mean and maximum duration) during the growing season are critical for the estimation of the final yield and damages on the crop. Finally indication of the total rainfall, in terms of number of rainy days and total precipitation amounts, has also been examined.

ID	*Index*	*Definition*
A	ONSET GS	Day (after 1 August) occurring after at least 40mm rain in 4 days
B	CESSATION GS	Day (after 1 February) occurring after 3 consecutive decads experiencing each < 20mm
C	LENGTH GS	Duration of growing season
D	MAXIMUM DRY SPELL	Length of maximum dry spell in the growing season (number of dry days)
E	MEAN DRY SPELL	Length of mean dry spell in the growing season
F	FRACTION OF RAINY DAYS	Fraction of rainy days within the growing season
G	TOTAL PRECIPITATION	Total precipitation during the growing season

TABLE 5.1: Indices used to investigate the rainfall characteristics during the growing season.

5.5.1 Indices for a 20^{th} century validation period

The next Figures, from 5.17 to 5.23, show the quantiles of the distributions of simulated indices computed separately for eleven growing seasons in the validation period, from 1985 to 1995, and the produced indices from the recorded rainfall according to the definitions given in Table 5.1. The 'simulated' indices have been derived from the 9-site average daily time series obtained from the previously simulated rainfall synthetic sequences (Section 5.4).

The first two indices represent the day of the year (from 1 to 365 or 366 for leap years) identifying the onset and cessation of the growing season. The year-to-year variability of such indices can be extremely high, especially for the onset, varying by as much as 3 months. However, the simulated indices seem to reproduce such variability pretty well, with a couple of years exception for the onset (1991 – 1993). The same is true for the growing season length (see Figure 5.19). The growing season cessation appears to be the best reproduced index (see Figure 5.18).

The range of the simulated length of the maximum dry spell within the growing season (Figure 5.20) encompasses the observed variability. However, the simulations tend to underestimate the mean dry spell length during the growing season, as seen in Figure 5.21.

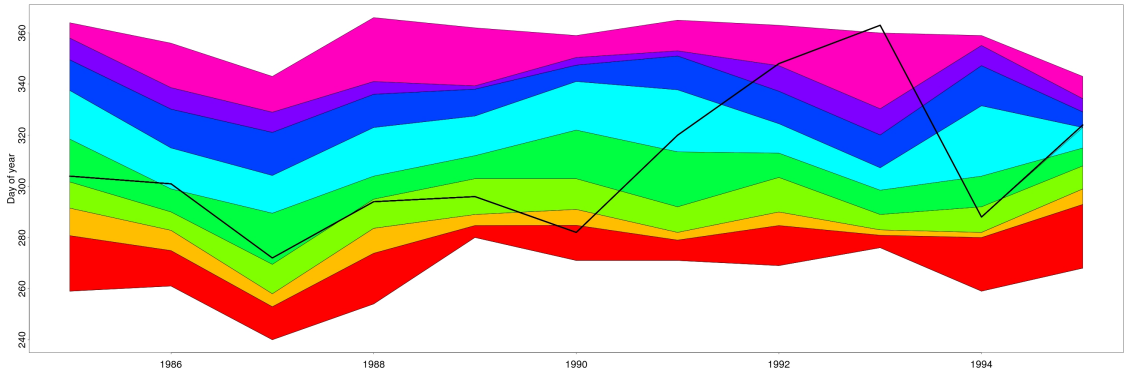


FIGURE 5.17: INDEX A – Observed and simulated onset of the growing season (day of the year) computed from rainfall spatial averages over 9 selected sites. The coloured bands represent the range of the distribution and the 0-5th, 5th-10th, 10th-25th, 25th-50th, 50th-75th, 75th-90th, 90th-95th and 95th-100th percentiles of 100 simulations. Thick black lines represent the 5th–95th percentile envelope of 100 imputations.

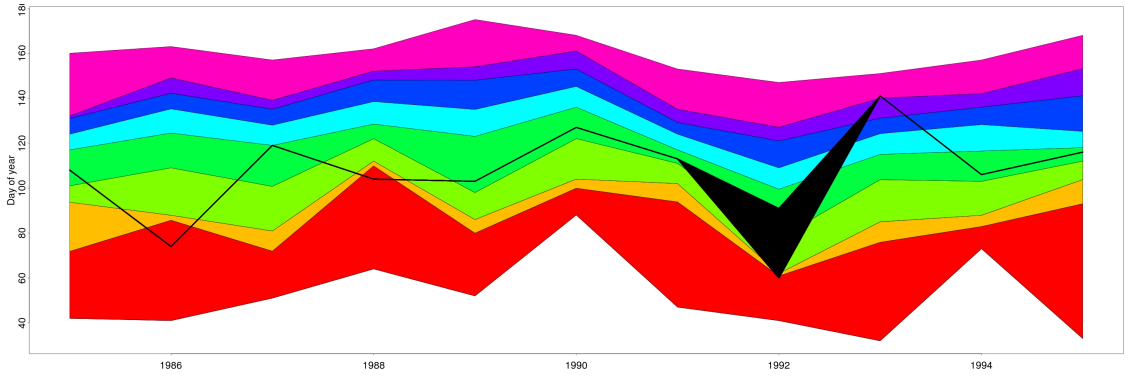


FIGURE 5.18: INDEX B – Observed and simulated cessation of the growing season (day of the year) computed from rainfall spatial averages over 9 selected sites. The coloured bands represent the range of the distribution and the 0-5th, 5th-10th, 10th-25th, 25th-50th, 50th-75th, 75th-90th, 90th-95th and 95th-100th percentiles of 100 simulations. Thick black lines represent the 5th–95th percentile envelope of 100 imputations.

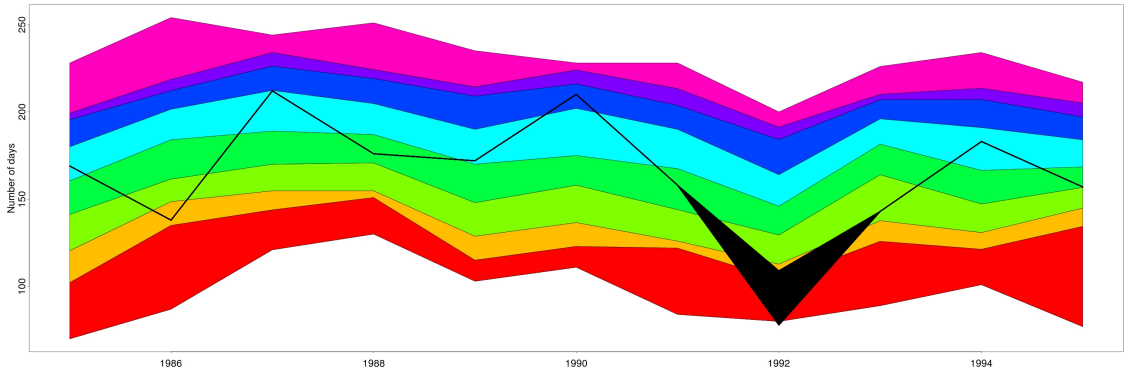


FIGURE 5.19: INDEX C – Observed and simulated length of the growing season (number of days) computed from rainfall spatial averages over 9 selected sites. The coloured bands represent the range of the distribution and the 0-5th, 5th-10th, 10th-25th, 25th-50th, 50th-75th, 75th-90th, 90th-95th and 95th-100th percentiles of 100 simulations. Thick black lines represent the 5th–95th percentile envelope of 100 imputations.

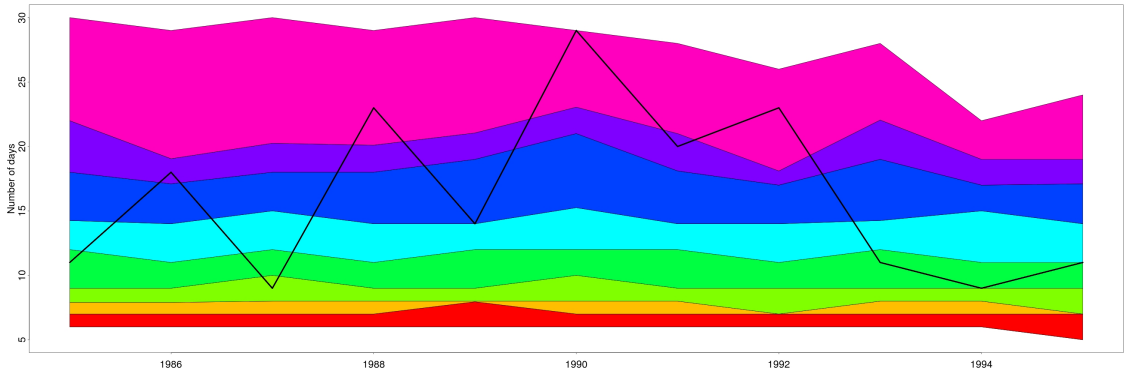


FIGURE 5.20: INDEX D – Observed and simulated maximum dry spell length within the growing season (number of days) computed from rainfall spatial averages over 9 selected sites. The coloured bands represent the range of the distribution and the 0-5th, 5th-10th, 10th-25th, 25th-50th, 50th-75th, 75th-90th, 90th-95th and 95th-100th percentiles of 100 simulations. Thick black lines represent the 5th–95th percentile envelope of 100 imputations.

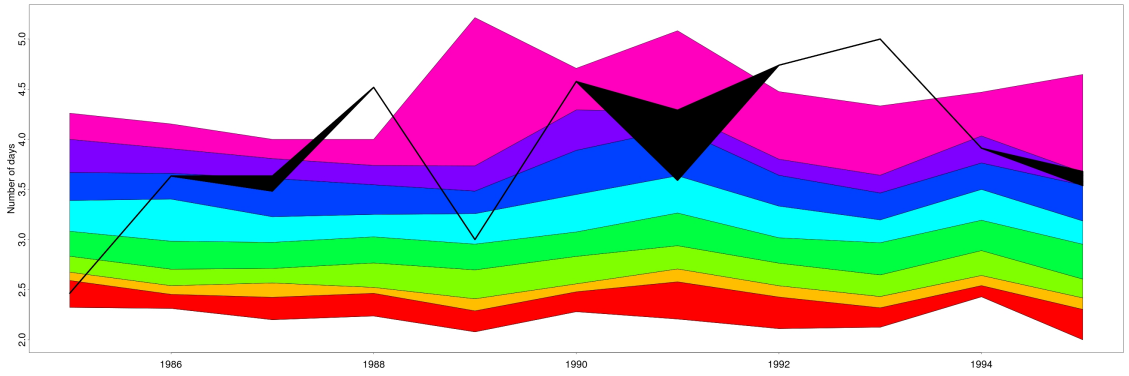


FIGURE 5.21: INDEX E – Observed and simulated mean dry spell length within the growing season (number of days) computed from rainfall spatial averages over 9 selected sites. The coloured bands represent the range of the distribution and the 0-5th, 5th-10th, 10th-25th, 25th-50th, 50th-75th, 75th-90th, 90th-95th and 95th-100th percentiles of 100 simulations. Thick black lines represent the 5th–95th percentile envelope of 100 imputations.

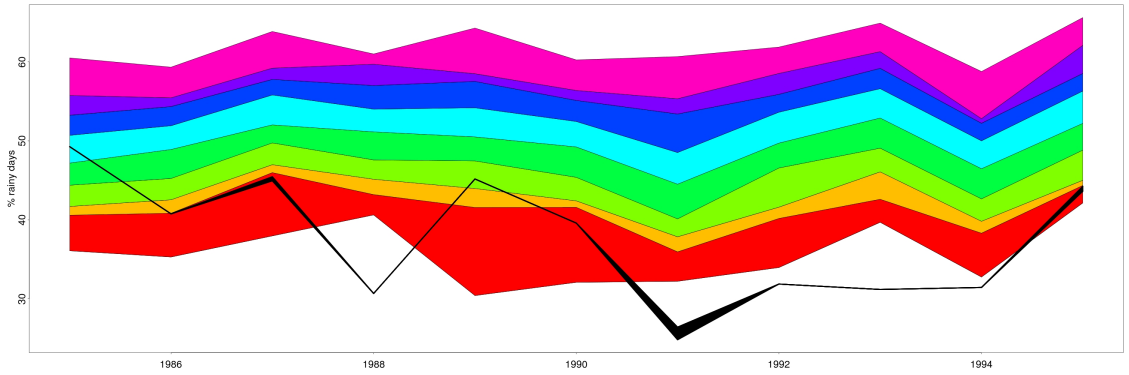


FIGURE 5.22: INDEX F – Observed and simulated fraction of rainy days during the growing season (%) computed from rainfall spatial averages over 9 selected sites. The coloured bands represent the range of the distribution and the 0-5th, 5th-10th, 10th-25th, 25th-50th, 50th-75th, 75th-90th, 90th-95th and 95th-100th percentiles of 100 simulations. Thick black lines represent the 5th–95th percentile envelope of 100 imputations.

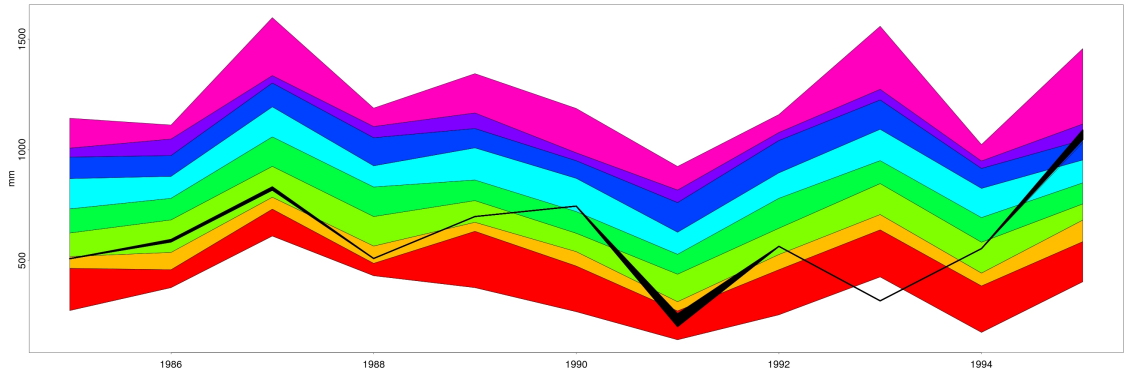
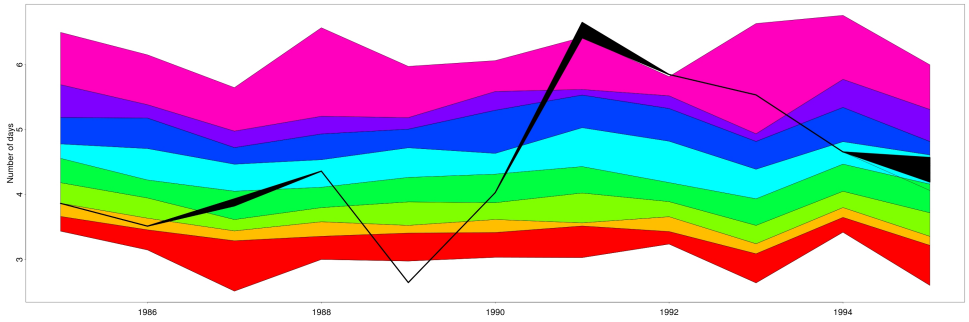


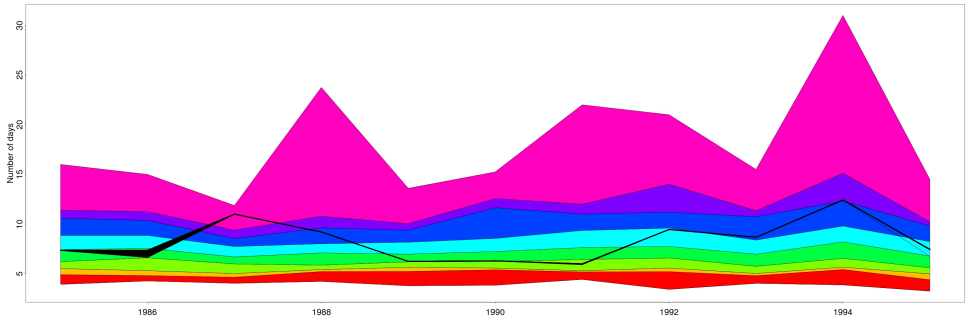
FIGURE 5.23: INDEX G – Observed and simulated total precipitation during the growing season (mm) computed from rainfall spatial averages over 9 selected sites. The coloured bands represent the range of the distribution and the 0-5th, 5th-10th, 10th-25th, 25th-50th, 50th-75th, 75th-90th, 90th-95th and 95th-100th percentiles of 100 simulations. Thick black lines represent the 5th–95th percentile envelope of 100 imputations.

The worst simulated individual index is the proportion of rainy days (Figure 5.22). Here the simulated values are clearly too high throughout the period investigated, in agreement with the results from Figure 5.16, top plot. Similarly, there, the occurrence of no rainy days was undersimulated, the problem being attributed to the spatial correlation in the logistic model. Indeed, if growing season indices for selected individual stations are investigated the problem appears to be reduced. Figures 5.24 and 5.25 show the plots for the most problematic indices: index E (mean dry spell length) and index F (fraction of rainy days), at 3 of the individual stations contributing to the 9-site average (station codes: S006, S035 and S042). For the individual stations there does not seem to be a bias in the distribution of simulations and, especially as far as the proportion of rainy days is concerned (index F), the results look much better.

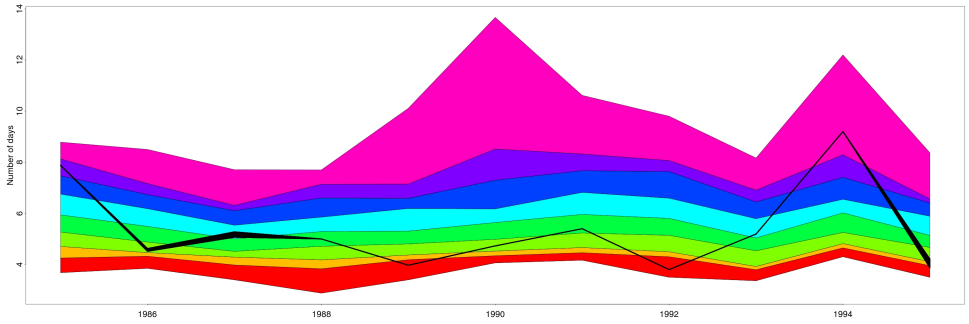
FIGURE 5.24: INDEX E – Same as Figure 5.21, but for selected individual stations.



(a) Station Code: S006



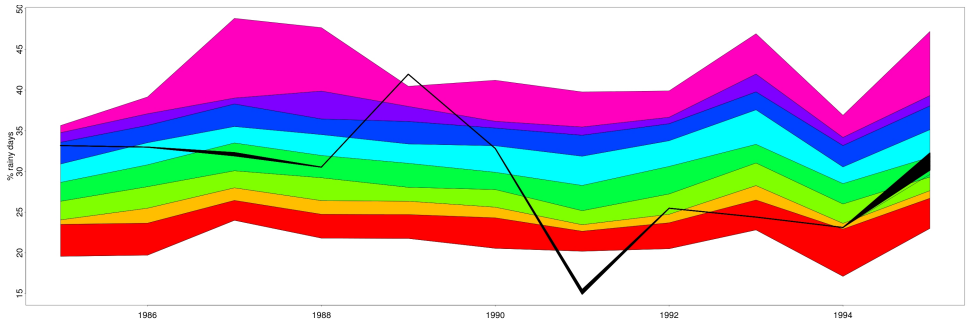
(b) Station Code: S035



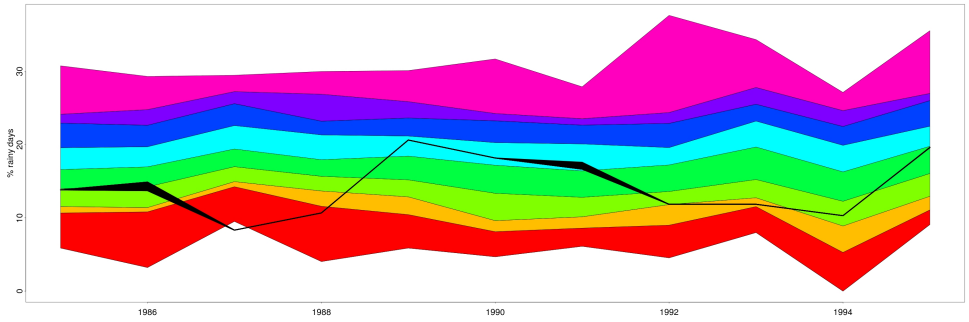
(c) Station Code: S042

Finally, although Figure 5.23 shows a slight tendency for the average simulated total precipitation (mm) during the growing season to be slightly higher than recorded, this tendency is not excessive; moreover, the simulated series capture the year-to-year variability and the range looks appropriate.

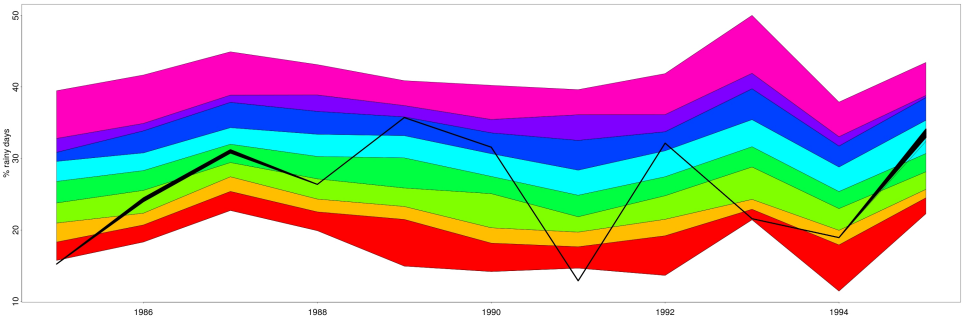
FIGURE 5.25: INDEX F – Same as Figure 5.22, but for selected individual stations.



(a) Station Code: S006



(b) Station Code: S035



(c) Station Code: S042

5.5.2 Projected precipitation and indices for a 20th century validation period

Having evaluated the ability of the GLM simulations to reproduce indices that are relevant for agricultural impacts studies, in this section a similar investigation is carried out for simulation driven by GCM outputs for the 20th century.

Recall from Section 3.2 that the climate indices used as covariates in this work have all been standardised. By standardising the outputs of each GCM in a similar way, any GCM-specific biases in the climate indices should be removed. That allows a direct comparison between the outcomes of the rainfall simulations and the subsequent derived growing season indices. As an example, Figure 5.26 shows the monthly distributions of precipitation for the 1985–1996 validation period driven by 6 GCMs compared with the observation and reanalysis driven distribution (far left). The distributions are essentially indistinguishable and thus the monthly standardisation is appropriate in order to remove GCM biases for rainfall simulation purposes. However, the statistical properties of the derived growing season indices do not appear to be perfectly standardised. Figures from 5.27 to 5.30 show the distributions of some of the growing season indices derived from the observations (extreme left boxplot) and from the GCM outputs, for the same period and stations. One reason may be the fact that the framework used during the fitting and rainfall simulation (GLM) is non linear. In such case a simple standardisation by mean and standard deviation does not appear to be enough to remove biases from the GCM outputs. A possible cause of the discrepancies in Figures 5.27 to 5.30 is that the persistence of anomalies in the climate indices is misrepresented in the GCMs. Under the GLM simulation methodology, this would lead to corresponding biases in the persistence of periods of above-or-below normal rainfall which, in turn, could affect indices such as the growing season onset. Indeed, the onset (Figure 5.27) and cessation (not shown here) of the growing season are placed a little later in the season, by on average around 5-15 and 3-5 days respectively. The range of length of the growing season is pretty well reproduced (see Figure 5.28), although the index mean appears a little reduced in few cases by a maximum of 10 days. Also well reproduced and within the bounds of the observed driven simulations are the maximum (not shown here) and mean (see Figure

5.29) dry spell length during the growing season. The means of the distribution of proportion of rainy days (Figure 5.30) as well as the total precipitation during the season (not shown here) tend to be slightly lower when the simulations are driven by almost all the selected GCMs, although the range of the distributions falls within the bounds of the observation driven indices.

The plotted results show the ability of the GLM framework in simulating credible precipitation sequences when driven by standardised GCMs outputs. However, minor biases still remain in the growing season indices simulations, suggesting that the computation of standard monthly anomalies may not be the best standardisation procedure when driving sophisticated statistical downscaling models.

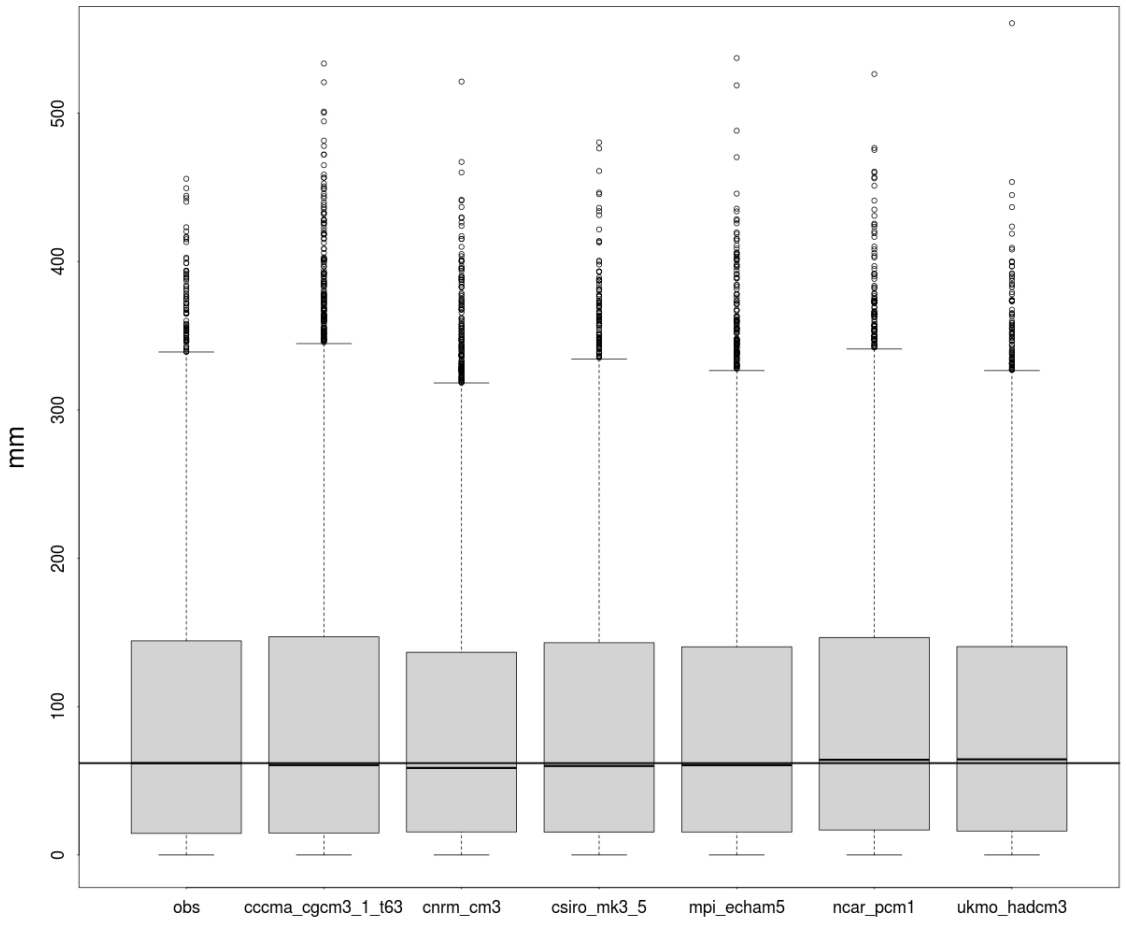


FIGURE 5.26: Distribution of simulated monthly precipitation totals, 1985–1996, for simulations driven by observations and reanalysis climate drivers (extreme left) and by six selected GCMs. The distributions are for the average rainfall over the 9 validation sites.

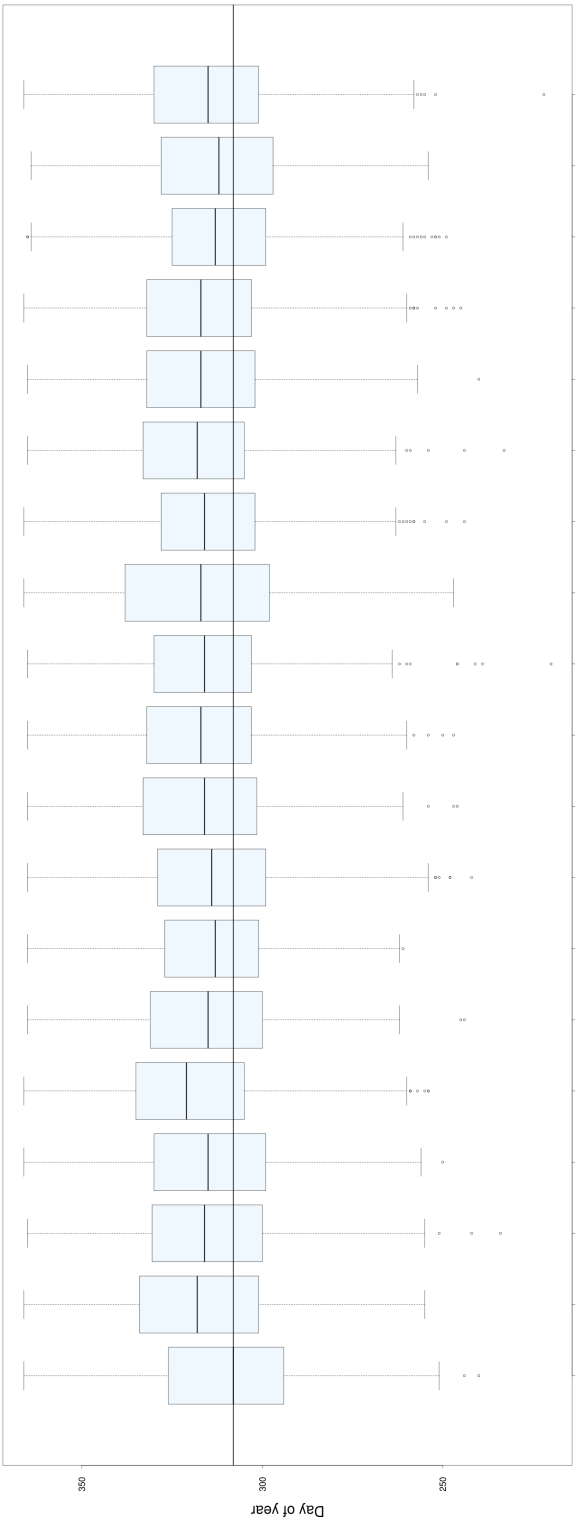


FIGURE 5.27: INDEX A – Distribution of 100 simulations of GS onset date driven by observational and GCM-derived climate indices for a 9-site average over the validation period 1985–1995.

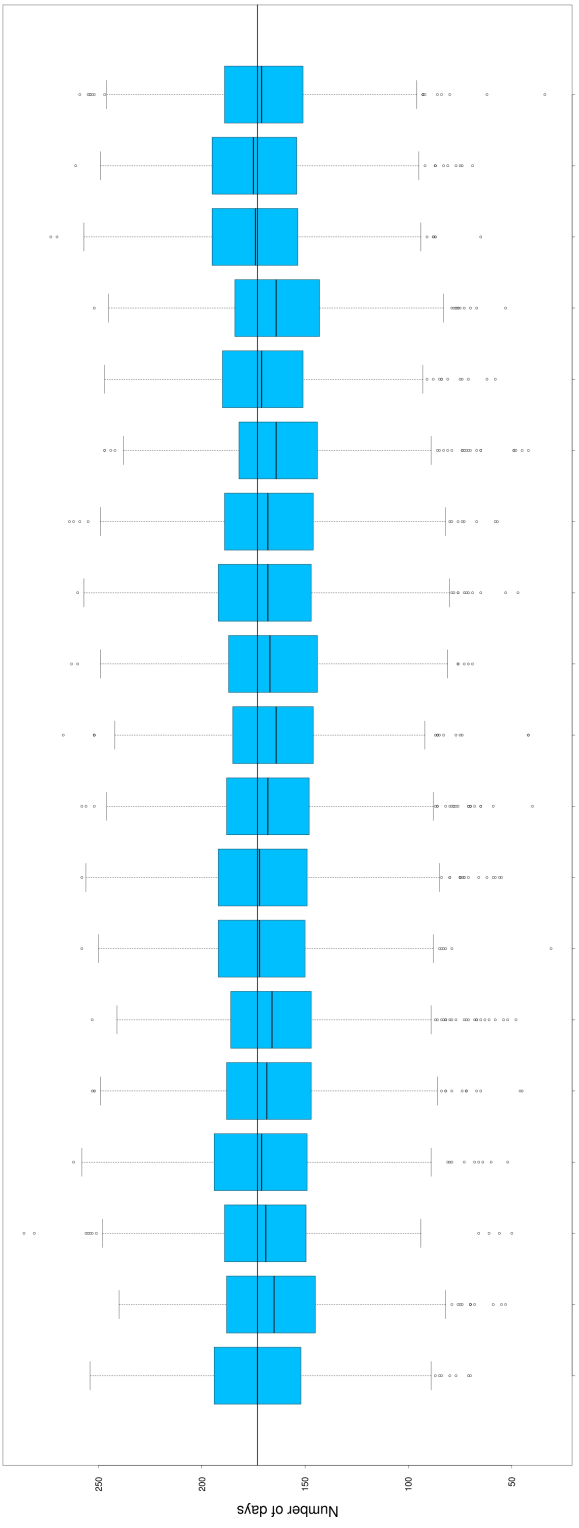


FIGURE 5.28: INDEX C – Distribution of 100 simulations of GS length driven by observational and GCM-derived climate indices for a 9-site average over the validation period 1985–1995.

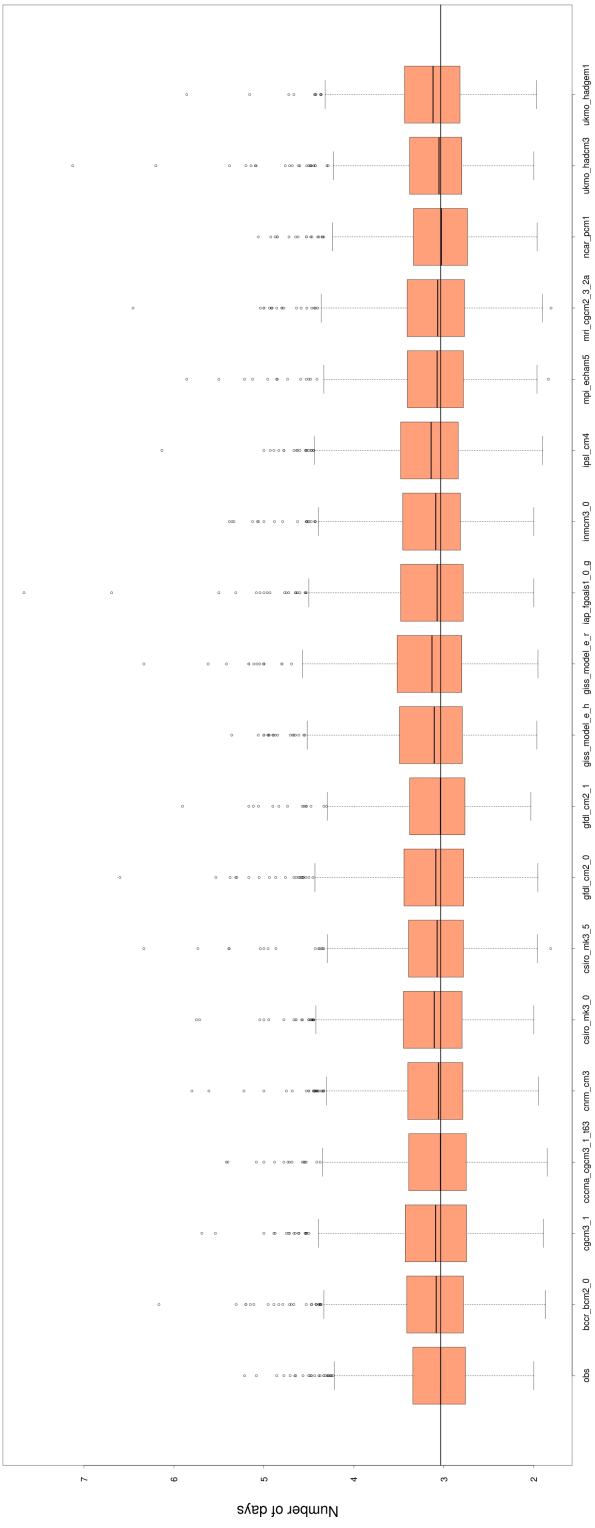


FIGURE 5.29: INDEX E – Distribution of 100 simulations of mean dry spell length during GS driven by observational and GCM-derived climate indices for a 9-site average over the validation period 1985–1995.

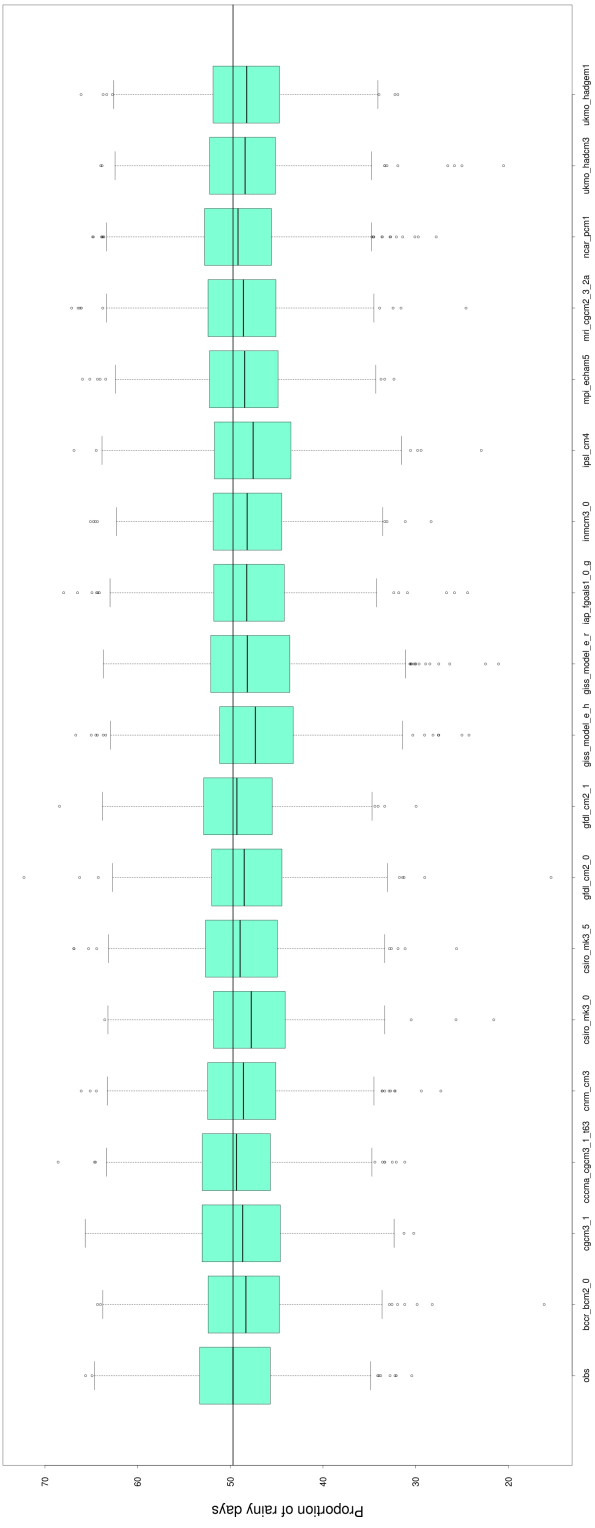


FIGURE 5.30: INDEX F – Distribution of 100 simulations of proportion of rainy days during GS driven by observational and GCM-derived climate indices for a 9-site average over the validation period 1985–1995.

5.5.3 Projected indices for a 21st century a1b scenario period

Finally, in order to provide useful information concerning the potential climate change impacts on future growing season characteristics for the local farmers, likely changes in the defined growing season indices, between the 20th and 21st centuries, conditional on GCM fields, have been examined. The purpose of such comparison is the investigation of the changes in the defined metrics based on 1 future scenario by the selected up-to-date climate models. The period 2026 – 2036 was chosen because considered a time period relevant for impact assessments and mitigation planning in agriculture (Lobell et al., 2008; Alston et al., 2000).

Figures 5.31 to 5.37 display the projected changes in distributions, between the *Control* (1985 – 1995; pale boxes) period and the *Future* (2026 – 2036, intermediate scenario a1b; dark boxes) period, of the growing season indices previously introduced (Section 5.5). For a direct comparison with the historical record the far left red boxplot depicts the distribution of the growing season indices for the available recorded period 1957 – 2000. Within the eighteen plotted distributions no consistent strong signal exists throughout the considered indices. The indices' internal variability was already high in the control period, a feature that may contribute in masking changes in the future projections. However, a few remarks can be made. The strongest changes in the median of the distributions are visible, for some of the GCMs, in the onset of the growing season as well as in the proportion of rainy days and total rainfall during the season (Figures 5.31, 5.36 and 5.37). Little change is visible in the expected maximum and mean length of dry spells (Figures 5.34 and 5.35).

Eleven out of the eighteen GCMs project a delay in the expected growing season onset. Such result is also confirmed by Shongwe et al. (2009), although only ten GCMs were considered in a study looking at the second half of the 21st century. On the other hand, a little change in the distribution appears to happen in the growing season cessation (Figure 5.32). Some of the GCMs project a slight earlier cessation in the mean index under the future scenario, increasing the risk for an early end of the same.

Although the change in projected growing season length distributions driven by GCMs is not homogeneous among the models, the medians of the distributions for each GCM

indicate a reduction in the same. The results shown in Figures 5.34 and 5.35 are consistent with what reported in Section 5.5.2 about the GCM deficiencies in representing persistence of dry conditions. Finally, conflicting projections are produced for the last two indices: proportion of rainy days and total precipitation during the growing season. As expected, all the GCM-derived distributions of the proportion of rainy days during the growing season are biased and shifted upwards when compared to the historical index. Such result is consistent with that shown in Figure 5.22. However, no clear signal is identifiable in the future changes in distributions.

A reduction in the expected value of the final index, and for two thirds of the GCMs, suggests a slight reduction in precipitation amounts throughout the growing season (Figure 5.37). However, nine models project an increase in the lower quartile of the index, also when compared to the historical values, suggesting a reduction in moderate drought occurrences.

Overall, the changes in the expected index values are tiny if compared with the internal variability, due to the natural fluctuations of the climate system (see Section 1.3.2; recall that here the indices are computed separately for each growing season rather than being aggregated measures). Internal variability proved to be a very important source of variability in studies of both rainfall and temperature projections for the future decades (Hawkins and Sutton, 2009; Northrop, 2010). According to Hawkins and Sutton (2009), in a study concerning regional mean surface temperature projections (using 15 GCMs, 3 emission scenarios and 1 ensemble run), for lead times of the next few decades, the dominant contributions are internal variability and model uncertainty. More specifically when looking at the sources of uncertainty for southern Africa and the next decade, the highest percentage of variance is explained by internal variability. Similar results were found in the work carried on by Northrop (2010) with the use of 24 GCMs, 3 scenarios and all available ensemble runs, where the most important sources of uncertainty in explaining 2020 – 2049 precipitation projections were found to be due to the chosen GCM and internal variability.

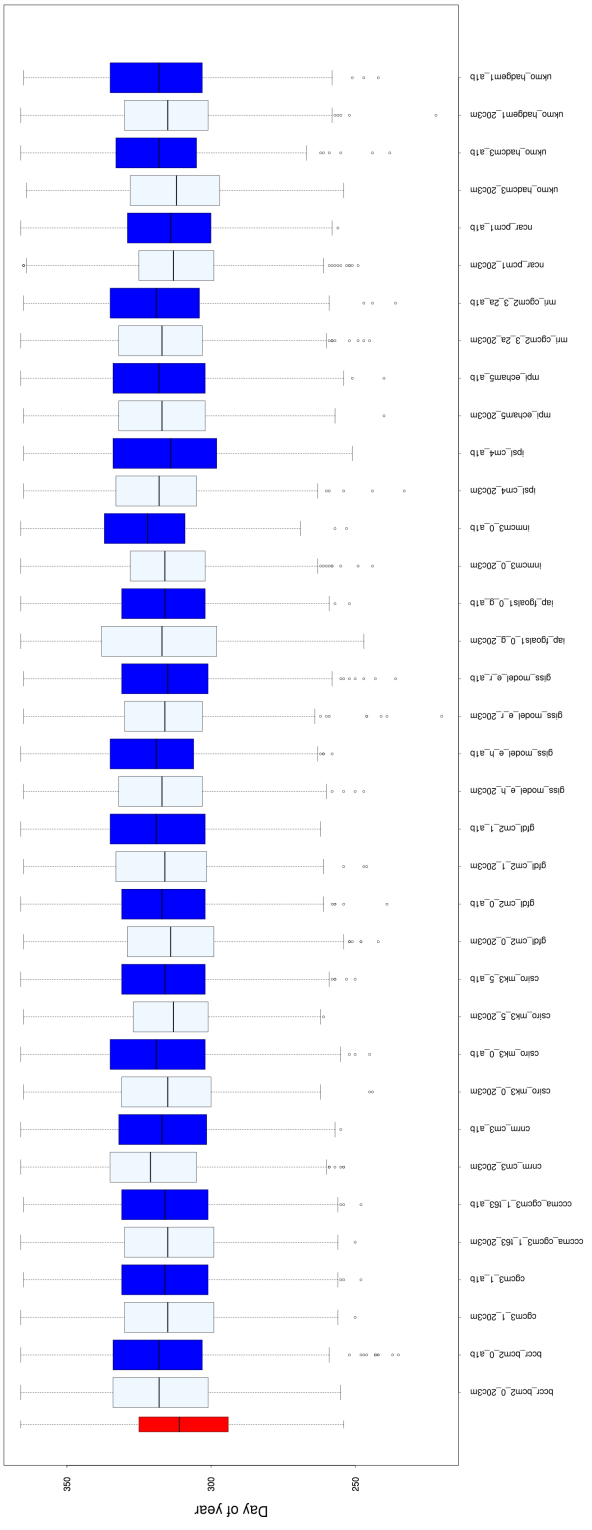


FIGURE 5.31: INDEX A – Changes in the distribution of 100 simulations of GS onset date driven by GCM derived climate indices compared with historical records (far left red boxplot). Pale boxes are the same as in Figures 5.26–5.30. Dark boxes are the corresponding simulations for the future.

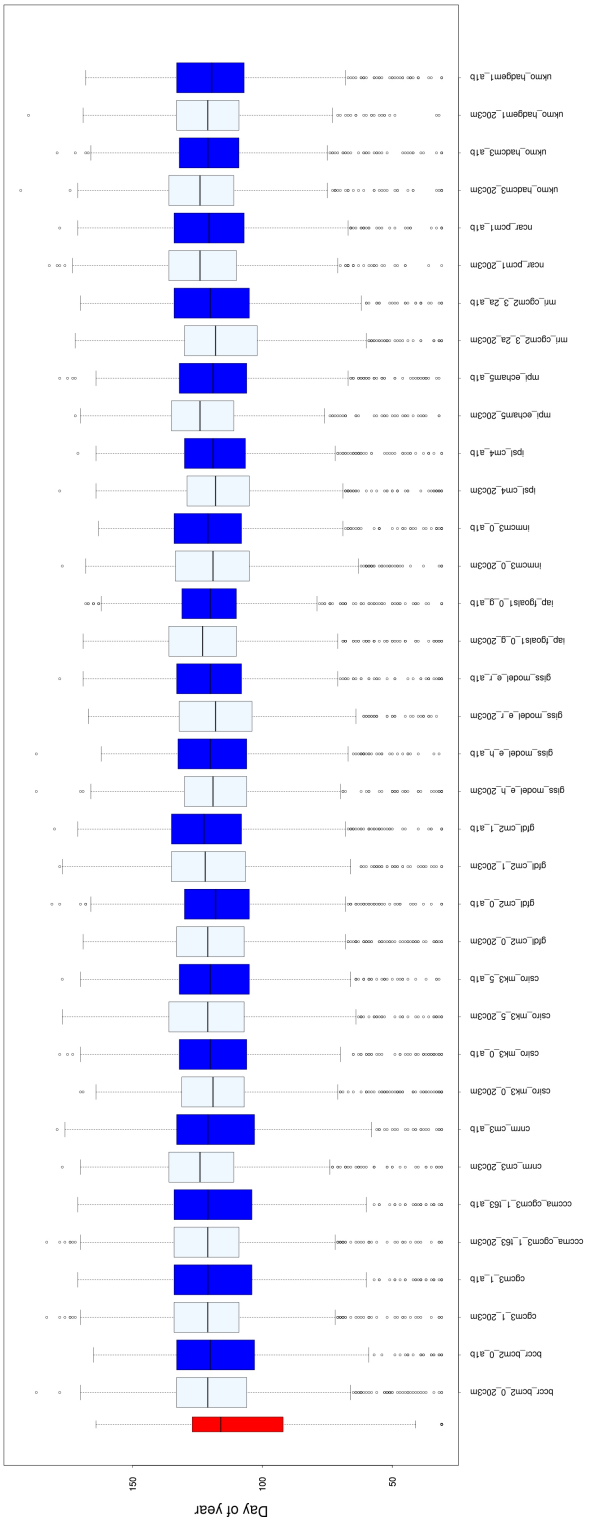


FIGURE 5.32: INDEX B – Changes in distribution of 100 simulations of GS cessation date driven by GCMs derived climate indices compared with historical records (far left red boxplot). Pale boxes are the same as in Figures 5.26–5.30. Dark boxes are the corresponding simulations for the future.

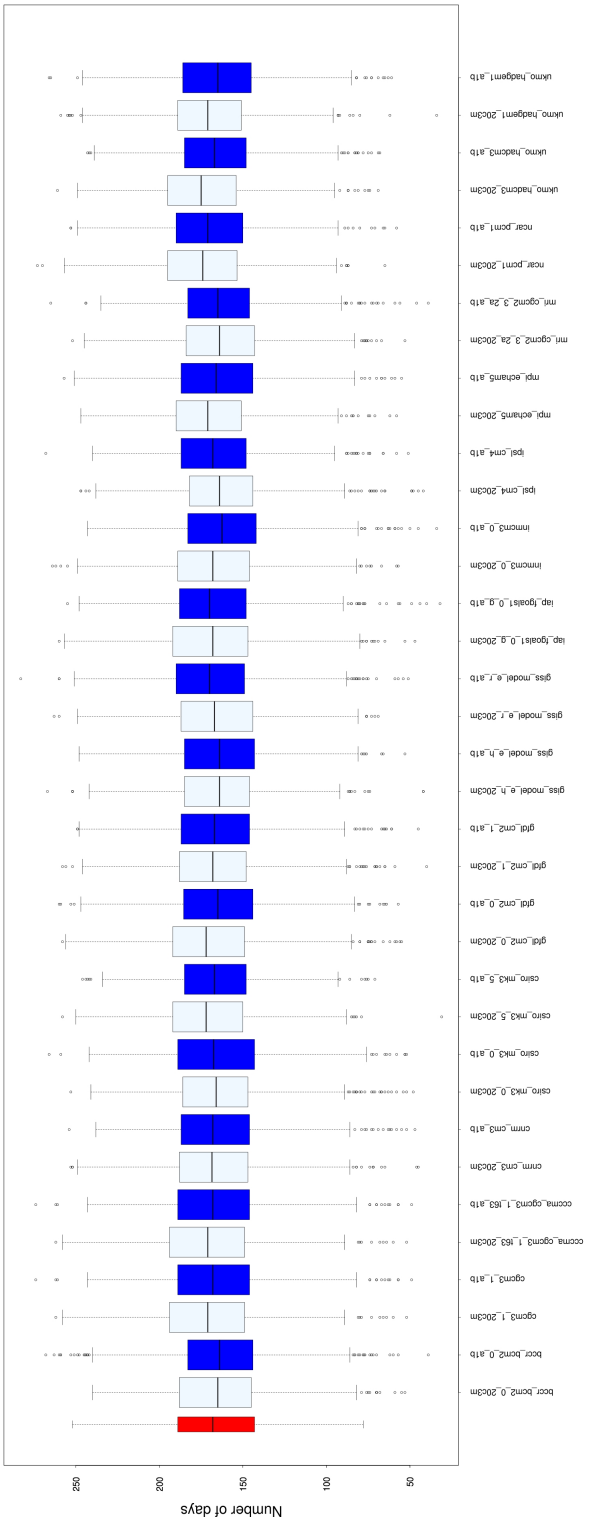


FIGURE 5.33: INDEX C – Changes in distribution of 100 simulations of GS length driven by GCMs derived climate indices compared with historical records (far left red boxplot). Pale boxes are the same as in Figures 5.26–5.30. Dark boxes are the corresponding simulations for the future.

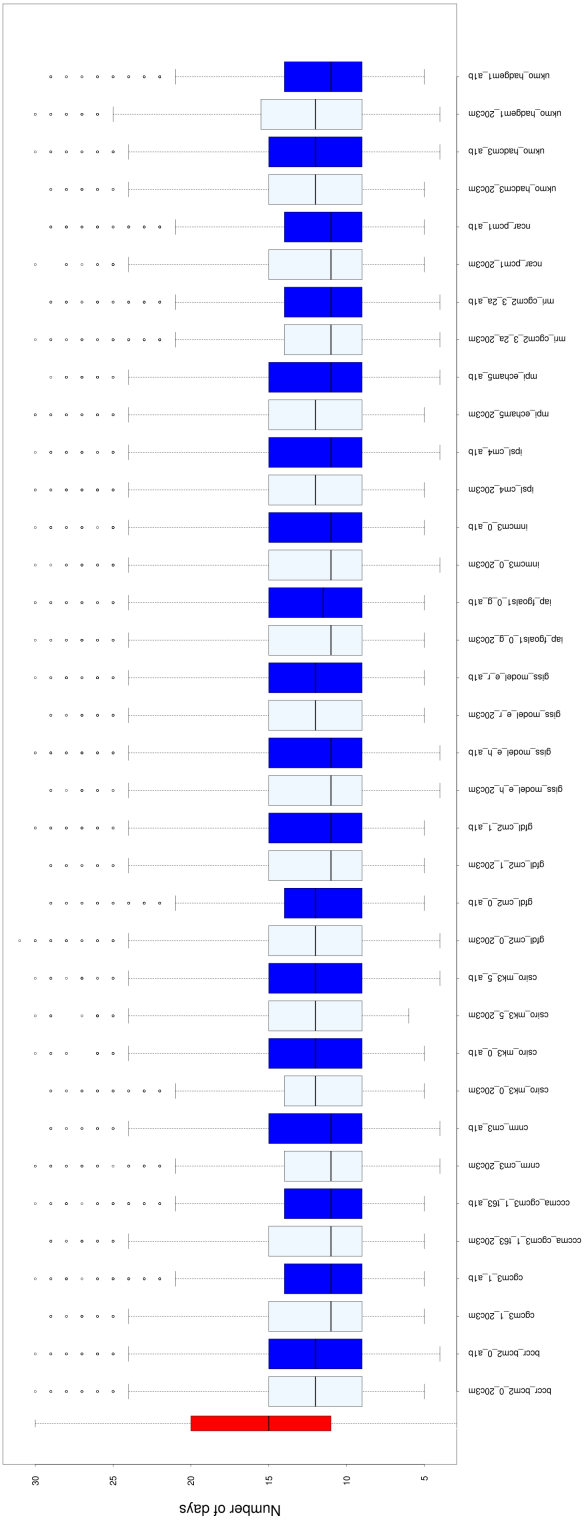


FIGURE 5.34: INDEX D – Changes in distribution of 100 simulations of maximum dry spell length during GS driven by GCMs derived climate indices compared with historical records (far left red boxplot). Pale boxes are the same as in Figures 5.26–5.30. Dark boxes are the corresponding simulations for the future.

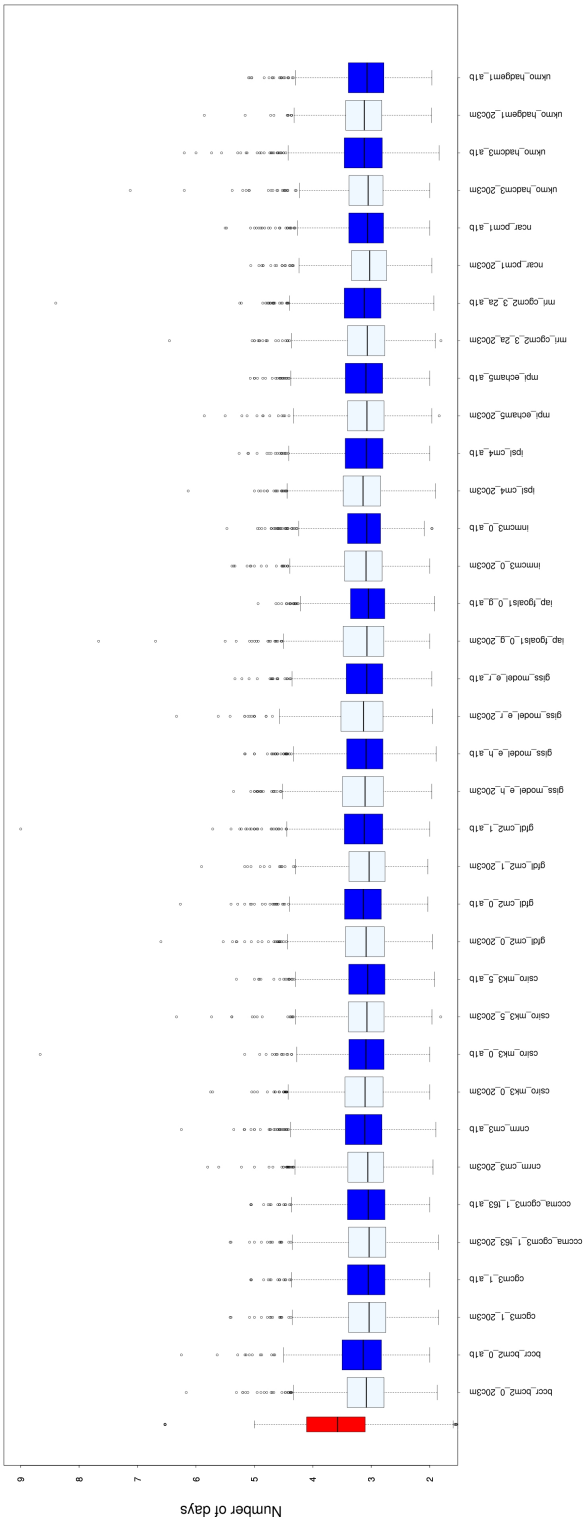


FIGURE 5.35: INDEX E – Changes in distribution of 100 simulations of mean dry spell length during GS driven by GCMs derived climate indices compared with historical records (far left red boxplot). Pale boxes are the same as in Figures 5.26–5.30. Dark boxes are the corresponding simulations for the future.

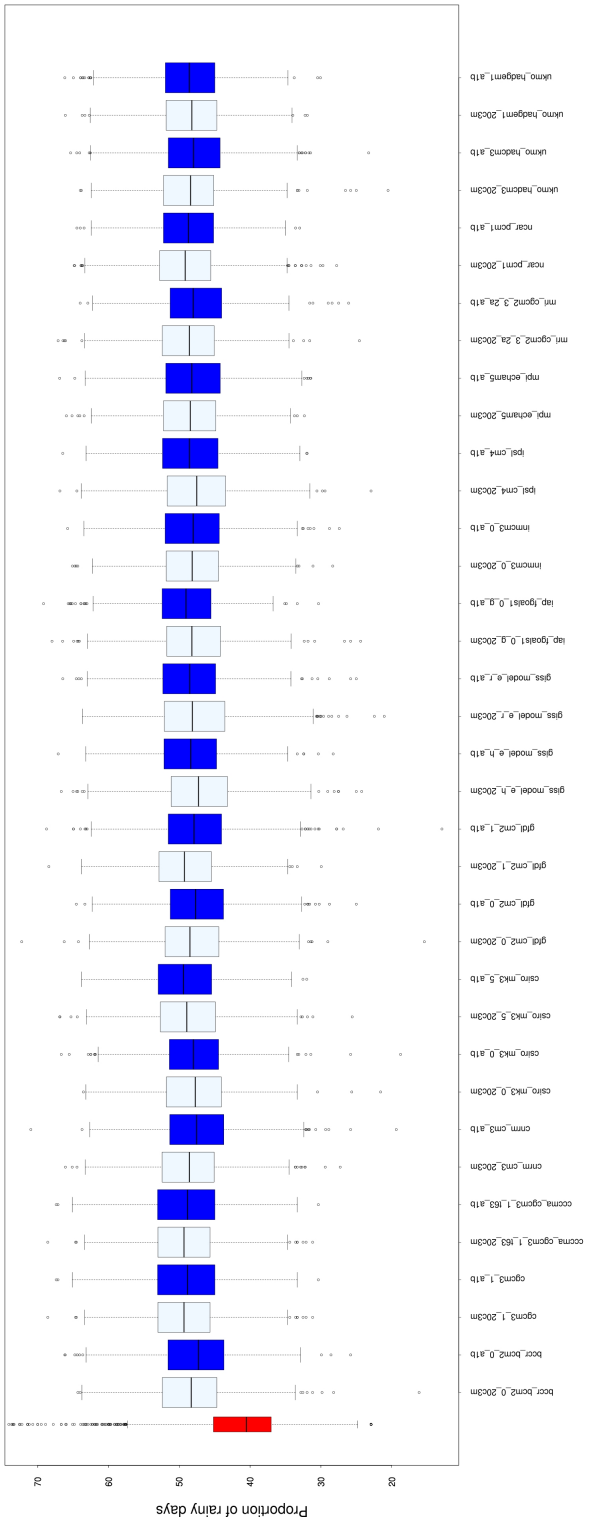


FIGURE 5.36: INDEX F – Changes in distribution of 100 simulations of proportion of rainy days during GS driven by GCMs derived climate indices compared with historical records (far left red boxplot). Pale boxes are the same as in Figures 5.26–5.30. Dark boxes are the corresponding simulations for the future.

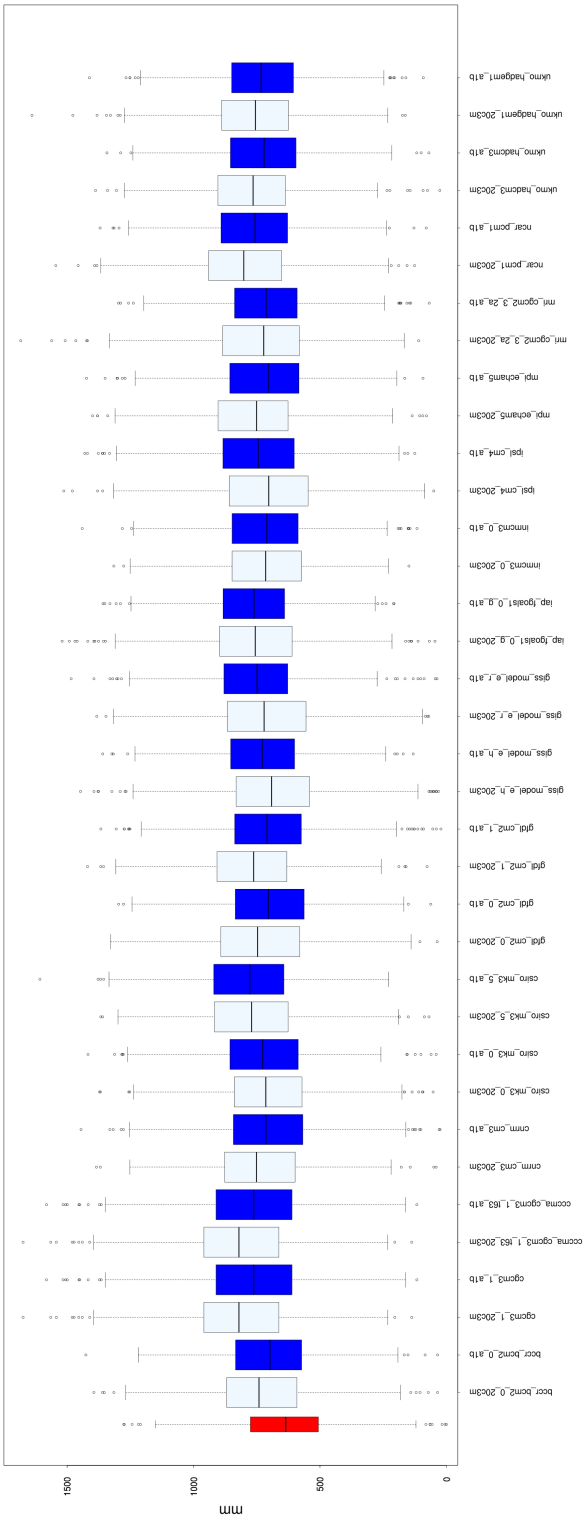


FIGURE 5.37: INDEX G – Changes in distribution of 100 simulations of total precipitation in mm during GS driven by GCMs derived climate indices compared with historical records (far left red boxplot). Pale boxes are the same as in Figures 5.26–5.30. Dark boxes are the corresponding simulations for the future.

5.6 Conclusion

In this chapter, the stochastic models previously developed at a subcontinental scale (Chapter 3) have been adapted to downscale large-scale climate drivers with the purpose of simulating daily rainfall and investigating intraseasonal rainfall characteristics in an orographically complex area in South Africa. This has been the first attempt to evaluate the suitability of a GLM framework for generating precipitation sequences for use in agricultural impact studies, as well as one of few applications of GLMs in Africa (along with Kenabatho et al., 2011 and Kigobe et al. 2011).

In general, the models capture adequately the seasonal and annual rainfall structure in the area. Although they contain a large number of terms to represent the spatial location, some limitations persist in resolving the complex geography of the region due to inhomogeneities present at neighbouring locations, particularly in the case of the occurrence model (Figure 5.8). The model is unable, here, unable to resolve large differences in rainfall recorded at neighbouring locations. One reason may be the variability in resolving the complex geography of the region, another may be associated with differences in recording practices (well know to cause problems with this kind of analysis; Yan et al. 2006). However, the simple addition of extra geographical terms only appears to increase the chances of overfitting without resolving such inconsistencies among adjacent stations.

Simulation performances for an independent period of time have been tested showing the abilities of the models in representing seasonal mean rainfall (Figures 5.14). However, the simulated rainfall values tend to be slightly too high throughout the seasons. When examining the proportion of rainy days for number of sites, the condition of 0 sites experiencing rain is clearly undersimulated (Figure 5.16). This is a clear indication that the used spatial correlation structure is not completely appropriate for such a complex region and future research could be undertaken in such direction.

In addition to the basic rainfall properties, 7 indices were used to explore the intraseasonal characteristics of daily rainfall under future climate changes for agricultural implications. The indices were computed from daily rainfall averaged across 9 validation stations in the study area. Although previous studies computed similar indices for single stations or grid cells (Raes et al., 2004; Kniveton et al., 2009; Tadross et al., 2009),

the choice of spatially-averaged indices was motivated by the greater relevance, for area impact studies, of being able to represent multi-site rainfall characteristics.

Broadly speaking, the indices are well reproduced by the statistical framework, although the tendency to oversimulate the proportion of rainy days is still present (see Figure 5.22). On the other hand, when single stations are considered (Figure 5.25) this bias is no longer clearly visible, suggesting a better performance when computing the indices for single stations rather then for area averages. This result indicates once more that the model's spatial structure could benefit from further investigation.

Although monthly standardisation of the climate drivers had been performed, the growing season indices derived from GCM outputs remain biased and sensitive to long term persistence structure (Figures 5.27 to 5.30). This suggests that in a non-linear framework, such as GLM, a simple monthly standardisation of GCM fields does not guarantee correct reproduction of the statistical properties in the derived indices. This is another potential area for further research.

Finally, GCM-driven simulations were used to infer future projected changes in the same growing season indices, assuming an a1b emission scenario. No coherent signal of change is obvious for any of the indices, suggesting that changes in the intraseasonal characteristics are quite uncertain in the area of study, results also confirmed by Tadross et al. (2007). Moreover, the dominant source of variation appears to be the natural variability. Indeed, Christensen et al. (2007) already pointed out how climate change and natural variability signals may not be easily distinguishable in the future projections. Similarly, Shongwe et al. (2009) found that internal decadal rainfall variability in the subcontinent is likely to mask any systematic changes in the total rainfall up to at least 2050, and, where systematic biases in the modelled (by GCMs) 20^{th} century climate exist, a wide range of changes in rainfall characteristics tends to be projected. Nonetheless, when considering together the simulations driven by all available GCMs (not shown here) a slight delay in the expected onset of the growing season, and a resultant reduction in the length of the season, have been found. Tadross et al. (2007) and Shongwe et al. (2009) came to similar conclusions, although considering a reduced number of GCMs (7 and 12 respectively), different future periods (2046 – 2065 and 2051 – 2200) and a different domain in southern Africa (104 stations lying between 26° – 10°S and 23° – 41°E and grid cells between 35° – 10°S and 20° – 42°E).

The findings of this case study may be helpful in addressing local future management practices and mitigation actions. Because internal rainfall variability may be the dominant signal in the nearest future decades, strategic planning should focus on developing systems which are robust to variability rather than changes in the climate system. Moreover, according to Washington et al. (2006), the management of a future changed climate may benefit from practices used to deal with the current climate variability.

Chapter 6

Conclusions and Future directions

6.1 Summary

The main focus of this study has been the investigation of southern African rainfall variability, at different temporal and spatial scales, and its implications for the agricultural sector.

The region has been affected by high precipitation variability for several decades (Mason and Jury, 1997), contributing in increasing its sensitivity to likely future changes in the climate system. Therefore, the ability to assess such variability and changes in rainfall characteristics is critical when planning adaptation and mitigation responses in relevant sectors.

Firstly, southern African precipitation was characterised in term of its climatology and variability using an observed gridded land surface precipitation product at a monthly time scale, and the ability of the GCMs in reproducing it later assessed. Dynamic climate models (e.g. GCMs and RCMs) projections for the current and future climate are widely used in the climate science. However, the computing costs of running these models severely limits the length of the time series and the number of runs produced by them (Brissette et al., 2007). In addition, up-to-date GCMs do not provide reliable information below around 200 km resolution (Meehl et al., 2007).

In this contest, most of the climate models proved to be able to reproduce 20^{th} century patterns and properties of the examined quantities in qualitative terms. However,

large differences still persist in quantitative terms. Notably, important deficiencies still remain in the simulation of clouds and tropical precipitation (Solomon et al., 2007). In particular, the annual precipitation trend is poorly reproduced by all the models, and no consistent pattern is detectable among them (see Chapter 2). Therefore, in order to investigate the rainfall variability, its climate controls and its projections under a future climate scenario, the use and development of a statistical model was explored as an alternative to the dynamical models.

Statistical modelling provides an easy-to-use and computationally inexpensive alternative for climate modelling and downscaling. Numerous stochastic models are available in the literature and the generation of synthetic rainfall sequences has been at the centre of interest particularly for hydrological and agricultural impact studies. See Wilks and Wilby (1999); Srikanthan and McMahon (2001); Maraun et al. (2010) for extensive reviews on different techniques.

The used GLM framework makes possible to condition its parameters (e.g. rainfall) on large scale weather predictors (Wilks and Wilby, 1999). As other statistical downscaling methods, a GLM is based on the assumption that large scale weather exhibits a strong influence on local scale weather (Maraun et al., 2010). Moreover it allows the simultaneous investigation of several components of the climate system and their effects upon the considered rainfall variability (Chandler, 2005). In this way, the effect of different climate indices upon precipitation, and the modulation of one factor on others, is investigated using a single statistical model. This has enabled a more realistic assessment of uncertainty than simply working with anomalies, as done in most previous statistical studies of the climate system.

Having identified links between the observed large-scale predictors and the observed local-scale rainfall, the developed model was subsequently used to downscale GCM outputs with the aim of studying the future rainfall projection for the region. However, the simulation performance of the model at a scale of a few grid cells was not considered adequate. Therefore, to downscale the large-scale drivers at local scale, and when precipitation at a daily time step is required, the location-specific application of the same methodology at a higher spatial and temporal resolution is necessary.

The final purpose of the study was the analysis of intraseasonal rainfall characteristics relevant for maize cropping and the provision of information useful to address farmers on

their likely changes under a future climate. This was achieved through the investigation of selected growing season indices and their changes under a climate change scenario. Previous studies computed similar indices for single stations or grid cells (Raes et al., 2004; Kniveton et al., 2009; Tadross et al., 2009). However, the choice here of spatially-averaged indices was motivated by the greater relevance, for area impact studies, of being able to represent multi-site rainfall characteristics.
The present study emerges as the first comprehensive assessment of factors affecting southern African rainfall variability. Similarly, it is the first systematic assessment of GLM performance for rainfall simulations for agricultural impact studies.

The key findings of this study are:

- The main time-varying climate factors affecting regional rainfall variability are El Niño and southern hemisphere mean temperature, as it is widely known in the literature. Moreover, the jet stream proved to have an effect in explaining part of the variability in the region. Such findings should provide useful means of narrowing down candidate climatic drivers in future investigations of rainfall variability in specific parts of southern Africa;

- The developed regional statistical model showed a good performance in the simulation of gridded monthly rainfall at subcontinental scale and better results when compared with the rainfall generated directly by the GCMs, which tend to be systematically biased, justifying its use for the investigation of future regional precipitation;

- Although at subcontinental scale some GCMs indicate a slight negative trend either between the present and future considered periods or within the future projected span of time, the signal is not consistent and no definitive statement can be made;

- Monthly standardisation of GCM fields did not proved to guarantee the correct reproduction of statistical properties in rainfall derived indices when using a complex, non-linear framework as GLM;

- When driven by 18 selected GCMs no coherent change signal is found in the simulated distributions of the growing season indices between a 20^{th} and a 21^{st}

century period. Moreover, when considering an individual growing season the dominant source of variation, for the next two decades, appears to be the natural variability. Therefore, strategic planning should focus on building systems that are robust to such variability, which may, in turn, prove useful for adaptation and mitigation measures under future changes in the climate.

6.2 Future directions

The characterisation and simulation of the rainfall variability at different temporal and spatial scales in the subcontinent could surely be further expanded, as well as its implications for a variety of sectors and locations investigated.
In particular, several questions and points arisen in the present study remain unanswered and would be fruitful topics for future research:

- The controls of multi-decadal fluctuations in southern African precipitation is still poorly understood. A few possible mechanisms have been investigated in the past literature and their influence has been taken into consideration in the present study. However, no clear driver of the interdecadal variability signal in the regional precipitation could be found. Inherent problems in exploring such low-frequency variability include the intrinsic difficulty in understanding low-frequency oscillations that may result from the integration of oceanic forcing even in the absence of definite physical mechanisms. On the other end, the length of available record may not be sufficiently long for the detection of such low-frequency processes. Further investigation of the processes that may be involved and expansion on the data used may be advisable;

- The addition of the subtropical jet stream as driver of part of southern African rainfall variability has been fairly new and no previous attempt has been done in the development of an index representing such climate feature. Meticulous analyses have been performed in order to build a covariate able to describe the complex process, involving the combination of the jet stream intensity and location, affecting the rainfall variability in the region. However, problems arose during the

rainfall simulations driven by the GCM-derived indices. Although, the issues were due to the inability by some of the GCMs to correctly reproduce the location of the upper-level winds, attempts, which did not resolve the problem, were made to overcome the problem. Moreover, although the climate indices were all standardized, that operation did not proved to be sufficient to remove internal GCM biases. This problem may lay in the inaccuracy of the GCM in reproducing persistence of periods of above-or-below normal precipitation.

The choice of climate predictors that are satisfyingly reproduced by the GCMs proved to be extremely relevant in the research presented here. However, this tack may be challenging when considering complex indices representing specific atmospheric features. Future investigation on the definition of such indices may be done and possible further research may focus on the development of alternative approaches to post-process GCM outputs and correct deficiencies in situations such as these.

- The development of a spatial dependence structure in the occurrence model is complex. A new approach has been used where latent Gaussian sets of variables, correlated matching the probability of rain modelled by the occurrence model, are generated. However, such methodology is still experimental and further investigation is advisable. The current structure represents the occurrence of rainfall as clusters of convective processes decaying with distance. However, the number of dry days is undersimulated suggesting there may be days when the convection is inhibit throughout the region. One direction may focus on the development of a statistical model able to represent the two states: conditions favourable to convecting process leading to clusters of precipitation and conditions inhibiting convection throughout the region.

Appendix A

Rainfall Models

RAINFALL, OCCURRENCE MODEL

Coefficients:	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	16.977686	3.853337	4.406	1.05e-05	***
LatLeg1	-23.173260	6.869980	-3.373	0.000743	***
LongLeg1	29.855262	7.891605	3.783	0.000155	***
AltLeg1	10.496327	3.938581	2.665	0.007699	**
C1	7.594401	4.460346	1.703	0.088634	.
S1	7.944402	4.184520	1.899	0.057627	.
C2	-5.260460	1.862733	-2.824	0.004742	**
S2	0.996769	2.749467	0.363	0.716955	
Lag1WetMonth	0.639557	0.605779	1.056	0.291078	
LatLeg2	-10.461187	5.384309	-1.943	0.052028	.
LongLeg2	22.280860	7.217235	3.087	0.002021	**
LatLeg3	-7.198004	2.675237	-2.691	0.007132	**
LongLeg3	5.799792	3.461549	1.675	0.093838	.
C3	-0.398315	0.292343	-1.362	0.173043	
S3	0.858124	0.414908	2.068	0.038619	*
AltLeg2	1.037586	0.641784	1.617	0.105938	
RHU1.R	-0.035049	0.044908	-0.780	0.435129	
nino3.4	-0.021550	0.125766	-0.171	0.863950	
sht.stdanom	-0.295643	0.093343	-3.167	0.001539	**
jet.intensity.stdanom	0.114255	0.077100	1.482	0.138364	
jet.position.effect	2.668269	0.733744	3.637	0.000276	***
LatLeg1:LongLeg1	-32.072917	10.413231	-3.080	0.002070	**
AltLeg1:S1	3.336009	3.514977	0.949	0.342578	
AltLeg1:C1	5.276815	4.472450	1.180	0.238060	
AltLeg1:S2	2.757424	2.393010	1.152	0.249205	
AltLeg1:C2	0.660887	1.567822	0.422	0.673367	

LatLeg1:S1	-0.861287	7.037904	-0.122	0.902599	
LatLeg1:C1	-8.771433	8.231731	-1.066	0.286621	
LatLeg1:S2	-6.577352	4.499455	-1.462	0.143793	
LatLeg1:C2	-8.581649	3.007632	-2.853	0.004327	**
LongLeg1:S1	11.727236	8.347518	1.405	0.160058	
LongLeg1:C1	10.622586	9.027469	1.177	0.239317	
LongLeg1:S2	3.545305	5.352570	0.662	0.507743	
LongLeg1:C2	-1.949773	3.703685	-0.526	0.598582	
S1:LatLeg2	-9.970726	5.254600	-1.898	0.057759	.
C1:LatLeg2	-17.641766	6.641076	-2.656	0.007897	**
S2:LatLeg2	-9.391081	3.553037	-2.643	0.008215	**
C2:LatLeg2	-10.948925	2.682177	-4.082	4.46e-05	***
S1:LongLeg2	8.620782	7.242377	1.190	0.233919	
C1:LongLeg2	7.153056	8.286348	0.863	0.388009	
S2:LongLeg2	2.226366	4.527576	0.492	0.622907	
C2:LongLeg2	-3.867683	2.809403	-1.377	0.168607	
LatLeg1:AltLeg1	24.206444	6.285108	3.851	0.000117	***
LongLeg1:AltLeg1	14.405538	7.555364	1.907	0.056564	.
AltLeg1:LatLeg2	11.422483	3.907788	2.923	0.003467	**
AltLeg1:LongLeg2	7.086582	6.498214	1.091	0.275474	
C1:Lag1WetMonth	-0.277303	0.168312	-1.648	0.099445	.
S1:Lag1WetMonth	-0.089624	0.240957	-0.372	0.709930	
S2:Lag1WetMonth	0.001501	0.154368	0.010	0.992241	
C2:Lag1WetMonth	0.011484	0.164228	0.070	0.944252	
LatLeg1:Lag1WetMonth	1.642578	0.632547	2.597	0.009411	**
LongLeg1:Lag1WetMonth	-1.032231	0.752951	-1.371	0.170402	
Lag1WetMonth:LatLeg2	1.481823	1.017013	1.457	0.145107	
Lag1WetMonth:LongLeg2	-0.397335	0.718991	-0.553	0.580518	
AltLeg1:Lag1WetMonth	0.302442	0.258059	1.172	0.241203	
LongLeg1:LatLeg2	-17.214279	5.775232	-2.981	0.002876	**
LatLeg1:LongLeg2	-12.967178	6.514332	-1.991	0.046529	*
S1:LatLeg3	-3.253471	2.968663	-1.096	0.273106	
C1:LatLeg3	-6.149487	3.265497	-1.883	0.059677	.
S2:LatLeg3	-4.993132	2.250754	-2.218	0.026526	*
C2:LatLeg3	-9.518454	1.673887	-5.686	1.30e-08	***
AltLeg1:LatLeg3	8.623390	2.402461	3.589	0.000331	***
S1:LongLeg3	1.125368	3.475969	0.324	0.746122	
C1:LongLeg3	0.266699	4.044767	0.066	0.947428	
S2:LongLeg3	-2.991204	2.384772	-1.254	0.209735	
C2:LongLeg3	-3.027455	1.620997	-1.868	0.061811	.
AltLeg1:LongLeg3	6.728907	1.856956	3.624	0.000291	***
LatLeg1:S3	1.727549	0.502661	3.437	0.000589	***
LongLeg1:S3	0.836154	0.520601	1.606	0.108245	
LatLeg1:C3	-0.489928	0.460036	-1.065	0.286886	
LongLeg1:C3	0.163404	0.382658	0.427	0.669361	
S1:AltLeg2	1.040715	0.525204	1.982	0.047530	*

C1:AltLeg2	-0.540621	0.671981	-0.805	0.421098	
LatLeg1:AltLeg2	3.795762	1.227820	3.091	0.001992	**
LongLeg1:AltLeg2	0.550732	0.724414	0.760	0.447109	
LatLeg1:RHU1.R	0.226105	0.058435	3.869	0.000109	***
LongLeg1:RHU1.R	-0.050348	0.052810	-0.953	0.340399	
LatLeg2:RHU1.R	0.266848	0.071608	3.727	0.000194	***
LongLeg2:RHU1.R	0.023219	0.051889	0.447	0.654530	
S1:RHU1.R	-0.029949	0.017452	-1.716	0.086148	.
C1:RHU1.R	0.003499	0.018274	0.191	0.848139	
LatLeg1:nino3.4	0.508226	0.278888	1.822	0.068405	.
LongLeg1:nino3.4	-0.472769	0.176873	-2.673	0.007519	**
S1:nino3.4	-0.245734	0.133213	-1.845	0.065085	.
C1:nino3.4	-0.040626	0.111542	-0.364	0.715690	
LatLeg1:sht.stdanom	0.036265	0.172562	0.210	0.833548	
LongLeg1:sht.stdanom	-0.188193	0.125267	-1.502	0.133012	
S1:sht.stdanom	-0.056486	0.097266	-0.581	0.561415	
C1:sht.stdanom	0.083709	0.086417	0.969	0.332710	
jet.intensity.stdanom:jet.position.effect	0.880688	0.930506	0.946	0.343913	
S1:jet.intensity.stdanom	0.069968	0.097980	0.714	0.475161	
C1:jet.intensity.stdanom	0.033195	0.085799	0.387	0.698839	
RHU1.R:jet.intensity.stdanom	0.010089	0.010242	0.985	0.324587	
sht.stdanom:jet.intensity.stdanom	-0.100467	0.054158	-1.855	0.063587	.
S1:jet.position.effect	0.392624	0.328077	1.197	0.231408	
C1:jet.position.effect	2.771392	0.786554	3.523	0.000426	***
RHU1.R:jet.position.effect	-0.010924	0.033713	-0.324	0.745919	
sht.stdanom:jet.position.effect	0.196317	0.084053	2.336	0.019511	*
LatLeg1:LongLeg1:S1	-0.397676	10.443030	-0.038	0.969624	
LatLeg1:LongLeg1:C1	-16.283664	11.798118	-1.380	0.167528	
LatLeg1:LongLeg1:S2	2.735425	7.244247	0.378	0.705728	
LatLeg1:LongLeg1:C2	12.733262	4.677853	2.722	0.006488	**
LatLeg1:LongLeg1:AltLeg1	7.962417	7.949620	1.002	0.316532	
LatLeg1:LongLeg1:Lag1WetMonth	0.649604	1.515786	0.429	0.668244	
LatLeg1:AltLeg1:S1	15.491805	4.404961	3.517	0.000437	***
LatLeg1:AltLeg1:C1	11.052943	4.937338	2.239	0.025179	*
LatLeg1:AltLeg1:S2	11.683839	2.900071	4.029	5.61e-05	***
LatLeg1:AltLeg1:C2	2.549849	1.919383	1.328	0.184022	
LongLeg1:AltLeg1:S1	-3.172224	6.766332	-0.469	0.639195	
LongLeg1:AltLeg1:C1	-3.742135	8.367419	-0.447	0.654711	
LongLeg1:AltLeg1:S2	-6.933016	4.415094	-1.570	0.116346	
LongLeg1:AltLeg1:C2	-3.636493	2.717140	-1.338	0.180781	
AltLeg1:S1:LatLeg2	8.011564	3.323038	2.411	0.015913	*
AltLeg1:C1:LatLeg2	9.320376	3.588528	2.597	0.009397	**
AltLeg1:S2:LatLeg2	8.951254	2.405219	3.722	0.000198	***
AltLeg1:C2:LatLeg2	3.692411	2.052700	1.799	0.072049	.
AltLeg1:S1:LongLeg2	-1.315436	5.350089	-0.246	0.805781	
AltLeg1:C1:LongLeg2	-2.134765	6.917432	-0.309	0.757621	

```
AltLeg1:S2:LongLeg2                                    -3.886895   3.660920 -1.062 0.288360
AltLeg1:C2:LongLeg2                                    -2.280407   2.349919 -0.970 0.331837
LongLeg1:S1:LatLeg2                                    -9.020505   6.308785 -1.430 0.152765
LongLeg1:C1:LatLeg2                                   -18.318264   7.597008 -2.411 0.015898 *
LongLeg1:S2:LatLeg2                                    -9.459308   4.727779 -2.001 0.045415 *
LongLeg1:C2:LatLeg2                                    -2.770836   3.964950 -0.699 0.484657
LongLeg1:AltLeg1:LatLeg2                               14.538201   5.013098  2.900 0.003731 **
LatLeg1:S1:LongLeg2                                     5.033069   6.532942  0.770 0.441054
LatLeg1:C1:LongLeg2                                    -3.961553   7.442999 -0.532 0.594551
LatLeg1:S2:LongLeg2                                    -3.055632   4.916699 -0.621 0.534284
LatLeg1:C2:LongLeg2                                     2.596549   3.373055  0.770 0.441423
LatLeg1:AltLeg1:LongLeg2                                3.612706   4.792783  0.754 0.450981
LatLeg1:LongLeg1:RHU1.R                                 0.191747   0.107295  1.787 0.073920 .
S1:jet.intensity.stdanom:jet.position.effect            0.333770   0.361109  0.924 0.355334
C1:jet.intensity.stdanom:jet.position.effect            0.868778   0.987398  0.880 0.378931
RHU1.R:jet.intensity.stdanom:jet.position.effect        0.036209   0.039767  0.911 0.362544
sht.stdanom:jet.intensity.stdanom:jet.position.effect  -0.028584   0.107223 -0.267 0.789791
LatLeg1:LongLeg1:AltLeg1:S1                            16.933666   6.265016  2.703 0.006874 **
LatLeg1:LongLeg1:AltLeg1:C1                             0.316735   6.884868  0.046 0.963307
LatLeg1:LongLeg1:AltLeg1:S2                             2.131281   3.997309  0.533 0.593910
LatLeg1:LongLeg1:AltLeg1:C2                            -2.688395   2.564451 -1.048 0.294486
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
```

RAINFALL, AMOUNT MODEL

Coefficients:	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	2.003e+00	1.170e+00	1.711	0.087046	.
LatLeg1	1.116e+01	3.277e+00	3.407	0.000659	***
LongLeg1	-1.338e+01	3.887e+00	-3.442	0.000579	***
AltLeg1	1.532e+01	5.283e+00	2.900	0.003737	**
C1	2.578e+00	6.286e-01	4.100	4.13e-05	***
S1	1.275e+00	6.641e-01	1.920	0.054807	.
C2	-1.171e+00	4.985e-01	-2.349	0.018824	*
S2	-1.283e+00	5.588e-01	-2.297	0.021629	*
Loglag1	2.578e-02	8.393e-02	0.307	0.758770	
LatLeg2	-7.003e-01	1.842e+00	-0.380	0.703830	
LongLeg2	-6.083e+00	2.238e+00	-2.718	0.006563	**
Namib	3.640e-02	6.719e-02	0.542	0.587950	
LatLeg3	-1.246e+00	1.265e+00	-0.985	0.324804	
LongLeg3	-5.080e+00	1.352e+00	-3.757	0.000172	***
AltLeg2	9.928e+00	3.100e+00	3.202	0.001365	**
AltLeg3	3.752e-01	1.330e+00	0.282	0.777811	
LatLeg4	-1.339e-01	5.018e-01	-0.267	0.789540	
LongLeg4	-6.161e+00	1.036e+00	-5.945	2.79e-09	***
C3	1.169e-02	7.649e-02	0.153	0.878526	
S3	-3.214e-01	8.860e-02	-3.628	0.000286	***
RHU1.R	1.148e-02	6.844e-03	1.677	0.093488	.
nino3.4	8.577e-02	4.892e-02	1.753	0.079591	.
sht.stdanom	1.922e-02	3.513e-02	0.547	0.584234	
SAM	6.397e-02	2.033e-02	3.147	0.001649	**
swio	5.122e-02	3.389e-02	1.512	0.130624	
jet.intensity.stdanom	1.992e-03	1.385e-02	0.144	0.885625	
jet.position.effect	2.499e-01	6.225e-02	4.015	5.95e-05	***
LatLeg1:LongLeg1	-4.106e+00	4.330e+00	-0.948	0.343030	
AltLeg1:S1	6.283e-02	2.318e-01	0.271	0.786395	
AltLeg1:C1	5.209e-01	2.452e-01	2.125	0.033634	*
AltLeg1:S2	-1.605e-01	2.029e-01	-0.791	0.428894	
AltLeg1:C2	-2.934e-01	1.961e-01	-1.496	0.134567	
LatLeg1:S1	-1.523e+00	1.017e+00	-1.497	0.134337	
LatLeg1:C1	5.152e+00	1.094e+00	4.711	2.48e-06	***
LatLeg1:S2	-2.260e+00	7.856e-01	-2.876	0.004027	**
LatLeg1:C2	-4.565e+00	7.448e-01	-6.129	8.94e-10	***
LongLeg1:S1	3.971e+00	1.063e+00	3.734	0.000189	***
LongLeg1:C1	2.764e+00	1.156e+00	2.392	0.016783	*
LongLeg1:S2	-5.197e-01	9.314e-01	-0.558	0.576889	
LongLeg1:C2	5.060e-01	9.005e-01	0.562	0.574194	
S1:LatLeg2	3.851e-01	1.261e+00	0.305	0.759994	
C1:LatLeg2	1.439e+00	1.213e+00	1.187	0.235412	
S2:LatLeg2	-3.268e+00	1.012e+00	-3.229	0.001241	**

```
C2:LatLeg2                -1.626e+00  9.535e-01 -1.706 0.088107 .
S1:LongLeg2                1.077e+00  1.112e+00  0.968 0.333076
C1:LongLeg2                4.531e-01  1.205e+00  0.376 0.706953
S2:LongLeg2               -1.004e+00  1.015e+00 -0.989 0.322854
C2:LongLeg2                1.564e-01  9.312e-01  0.168 0.866661
LatLeg1:AltLeg1           -3.586e+01  1.157e+01 -3.099 0.001941 **
LongLeg1:AltLeg1           6.867e+01  1.382e+01  4.968 6.79e-07 ***
AltLeg1:LatLeg2           -3.673e+00  5.426e+00 -0.677 0.498453
AltLeg1:LongLeg2           4.328e+01  1.093e+01  3.960 7.51e-05 ***
LatLeg1:Loglag1            4.825e-02  2.634e-01  0.183 0.854642
LongLeg1:Loglag1          -2.262e-01  1.866e-01 -1.213 0.225297
Loglag1:LatLeg2           -7.858e-02  1.369e-01 -0.574 0.565989
Loglag1:LongLeg2          -5.029e-02  1.429e-01 -0.352 0.724858
C1:Loglag1                -2.812e-02  1.633e-02 -1.722 0.085038 .
S1:Loglag1                -1.364e-02  1.773e-02 -0.769 0.441671
C2:Loglag1                -2.367e-02  1.505e-02 -1.573 0.115775
S2:Loglag1                -7.307e-03  1.709e-02 -0.428 0.668912
AltLeg1:Loglag1            5.339e-02  2.665e-02  2.003 0.045137 *
S1:Namib                   3.311e-01  5.422e-02  6.107 1.02e-09 ***
C1:Namib                   2.513e-01  6.420e-02  3.914 9.11e-05 ***
LongLeg1:LatLeg2          -1.984e+01  4.855e+00 -4.087 4.38e-05 ***
LatLeg1:LongLeg2           2.872e+01  5.348e+00  5.371 7.90e-08 ***
S1:LatLeg3                -1.051e+00  7.186e-01 -1.462 0.143694
C1:LatLeg3                 2.101e+00  7.074e-01  2.969 0.002987 **
S2:LatLeg3                -1.722e+00  4.824e-01 -3.570 0.000358 ***
C2:LatLeg3                -2.983e+00  4.693e-01 -6.356 2.10e-10 ***
Loglag1:LatLeg3            1.275e-01  1.376e-01  0.926 0.354272
AltLeg1:LatLeg3           -1.989e+00  4.749e+00 -0.419 0.675321
S1:LongLeg3                3.758e-01  7.037e-01  0.534 0.593390
C1:LongLeg3                2.027e+00  7.724e-01  2.624 0.008683 **
S2:LongLeg3               -3.152e-01  6.205e-01 -0.508 0.611423
C2:LongLeg3               -1.103e+00  5.777e-01 -1.910 0.056184 .
Loglag1:LongLeg3          -7.173e-02  9.589e-02 -0.748 0.454420
AltLeg1:LongLeg3           3.841e+01  6.791e+00  5.655 1.57e-08 ***
S1:AltLeg2                 9.116e-02  3.893e-02  2.342 0.019195 *
C1:AltLeg2                 6.921e-03  4.097e-02  0.169 0.865874
S2:AltLeg2                 1.734e-02  3.059e-02  0.567 0.570848
C2:AltLeg2                 2.187e-02  3.223e-02  0.679 0.497410
LatLeg1:AltLeg2            1.045e+01  4.453e+00  2.348 0.018906 *
LongLeg1:AltLeg2           4.610e+00  7.504e+00  0.614 0.538965
LatLeg2:AltLeg2           -8.694e-01  2.010e+00 -0.432 0.665447
LongLeg2:AltLeg2           2.203e+01  7.165e+00  3.075 0.002105 **
LatLeg3:AltLeg2            9.490e-01  1.995e+00  0.476 0.634221
LongLeg3:AltLeg2           8.767e+00  4.278e+00  2.049 0.040459 *
Loglag1:AltLeg2           -3.189e-02  2.221e-02 -1.436 0.151017
S1:AltLeg3                -1.234e-01  4.126e-02 -2.991 0.002782 **
```

C1:AltLeg3	-1.840e-01	3.760e-02	-4.893	9.99e-07 ***
S2:AltLeg3	6.054e-02	2.946e-02	2.055	0.039873 *
C2:AltLeg3	8.827e-02	2.743e-02	3.218	0.001290 **
LatLeg1:AltLeg3	-1.213e+01	4.568e+00	-2.656	0.007915 **
LongLeg1:AltLeg3	1.655e+01	2.768e+00	5.981	2.24e-09 ***
LatLeg2:AltLeg3	-3.904e+00	2.151e+00	-1.815	0.069581 .
LongLeg2:AltLeg3	3.089e+00	9.231e-01	3.346	0.000821 ***
LatLeg3:AltLeg3	-4.916e+00	2.443e+00	-2.013	0.044150 *
LongLeg3:AltLeg3	7.428e+00	1.079e+00	6.883	5.98e-12 ***
Loglag1:AltLeg3	3.917e-02	1.859e-02	2.108	0.035055 *
LongLeg1:LatLeg3	-5.433e+00	1.027e+00	-5.291	1.22e-07 ***
LatLeg1:LongLeg3	1.014e+01	2.067e+00	4.905	9.37e-07 ***
LatLeg2:LongLeg2	-1.512e+00	1.531e+00	-0.988	0.323301
S1:LatLeg4	-5.915e-01	4.538e-01	-1.303	0.192472
C1:LatLeg4	5.503e-01	4.389e-01	1.254	0.209880
S2:LatLeg4	-1.029e+00	3.566e-01	-2.885	0.003913 **
C2:LatLeg4	-1.153e+00	3.521e-01	-3.274	0.001060 **
AltLeg1:LatLeg4	3.247e+00	1.758e+00	1.847	0.064703 .
S1:LongLeg4	-4.990e-01	3.648e-01	-1.368	0.171281
C1:LongLeg4	7.022e-01	3.646e-01	1.926	0.054114 .
S2:LongLeg4	-5.128e-02	2.956e-01	-0.173	0.862273
C2:LongLeg4	-4.564e-01	2.636e-01	-1.731	0.083427 .
AltLeg1:LongLeg4	2.132e+01	4.091e+00	5.212	1.88e-07 ***
Loglag1:S3	1.347e-02	1.278e-02	1.054	0.291835
AltLeg1:S3	-1.630e-02	2.738e-02	-0.595	0.551607
LatLeg1:S3	-2.707e-01	7.630e-02	-3.549	0.000388 ***
LongLeg1:S3	-1.015e-01	8.000e-02	-1.268	0.204724
LatLeg2:S3	-2.765e-01	1.029e-01	-2.688	0.007192 **
LongLeg2:S3	-9.272e-02	1.019e-01	-0.910	0.363010
Loglag1:C3	3.712e-02	1.136e-02	3.267	0.001087 **
AltLeg1:C3	4.198e-02	2.434e-02	1.724	0.084663 .
LatLeg1:C3	2.781e-01	6.776e-02	4.103	4.08e-05 ***
LongLeg1:C3	1.010e-02	7.153e-02	0.141	0.887772
LatLeg2:C3	2.975e-01	9.218e-02	3.228	0.001248 **
LongLeg2:C3	-2.815e-02	9.371e-02	-0.300	0.763855
LatLeg1:RHU1.R	2.155e-02	7.896e-03	2.729	0.006358 **
LongLeg1:RHU1.R	5.330e-02	7.939e-03	6.713	1.93e-11 ***
LatLeg2:RHU1.R	6.847e-02	1.090e-02	6.280	3.42e-10 ***
LongLeg2:RHU1.R	5.072e-02	1.103e-02	4.599	4.27e-06 ***
S1:RHU1.R	6.987e-03	3.851e-03	1.814	0.069613 .
C1:RHU1.R	9.275e-03	3.928e-03	2.361	0.018233 *
S2:RHU1.R	-2.359e-03	3.416e-03	-0.690	0.489910
C2:RHU1.R	6.277e-03	2.916e-03	2.153	0.031359 *
LatLeg1:nino3.4	1.258e-01	5.708e-02	2.204	0.027535 *
LongLeg1:nino3.4	-1.510e-01	5.910e-02	-2.555	0.010631 *
LatLeg2:nino3.4	1.673e-01	8.168e-02	2.048	0.040537 *

```
LongLeg2:nino3.4                                -6.598e-02  7.779e-02 -0.848 0.396365
S1:nino3.4                                      -1.913e-02  2.633e-02 -0.727 0.467532
C1:nino3.4                                      -9.548e-02  2.307e-02 -4.138 3.52e-05 ***
S2:nino3.4                                      -1.137e-02  2.308e-02 -0.493 0.622132
C2:nino3.4                                      -2.443e-03  2.238e-02 -0.109 0.913064
RHU1.R:nino3.4                                  -5.314e-03  1.660e-03 -3.201 0.001369 **
LatLeg1:sht.stdanom                              8.070e-02  4.242e-02  1.902 0.057147 .
LongLeg1:sht.stdanom                             1.723e-01  4.035e-02  4.270 1.96e-05 ***
LatLeg2:sht.stdanom                              1.885e-01  6.010e-02  3.136 0.001714 **
LongLeg2:sht.stdanom                             1.148e-01  5.613e-02  2.045 0.040841 *
S1:sht.stdanom                                   2.769e-02  1.710e-02  1.619 0.105387
C1:sht.stdanom                                   4.128e-02  1.529e-02  2.700 0.006930 **
S2:sht.stdanom                                   1.900e-02  1.636e-02  1.161 0.245577
C2:sht.stdanom                                   1.058e-03  1.488e-02  0.071 0.943343
RHU1.R:sht.stdanom                               6.588e-03  1.655e-03  3.980 6.90e-05 ***
LatLeg1:SAM                                     -7.227e-02  2.253e-02 -3.208 0.001337 **
LongLeg1:SAM                                     6.955e-02  2.248e-02  3.094 0.001975 **
LatLeg2:SAM                                      7.900e-02  3.583e-02  2.205 0.027454 *
LongLeg2:SAM                                     8.006e-02  3.189e-02  2.510 0.012063 *
S1:SAM                                           1.091e-02  8.933e-03  1.222 0.221820
C1:SAM                                           1.103e-02  9.758e-03  1.131 0.258231
LatLeg1:swio                                     5.704e-02  4.031e-02  1.415 0.157086
LongLeg1:swio                                   -3.163e-02  4.148e-02 -0.763 0.445721
LatLeg2:swio                                     9.800e-02  5.753e-02  1.703 0.088500 .
LongLeg2:swio                                    3.122e-03  5.455e-02  0.057 0.954367
S1:swio                                          6.293e-02  1.940e-02  3.244 0.001181 **
C1:swio                                         -2.969e-02  1.717e-02 -1.729 0.083801 .
S2:swio                                         -2.610e-02  1.661e-02 -1.571 0.116151
C2:swio                                          1.242e-02  1.679e-02  0.740 0.459361
jet.intensity.stdanom:jet.position.effect        7.696e-02  5.817e-02  1.323 0.185831
S1:jet.intensity.stdanom                         1.480e-02  1.738e-02  0.851 0.394560
C1:jet.intensity.stdanom                         6.127e-02  1.716e-02  3.571 0.000356 ***
SAM:jet.intensity.stdanom                        8.722e-04  6.091e-03  0.143 0.886136
sht.stdanom:jet.intensity.stdanom               -1.916e-02  9.984e-03 -1.919 0.054982 .
RHU1.R:jet.intensity.stdanom                     2.401e-03  1.798e-03  1.336 0.181718
S1:jet.position.effect                           4.193e-02  3.287e-02  1.276 0.202017
C1:jet.position.effect                           2.495e-01  7.573e-02  3.295 0.000986 ***
SAM:jet.position.effect                         -1.833e-02  5.609e-03 -3.268 0.001084 **
sht.stdanom:jet.position.effect                  1.852e-02  8.780e-03  2.109 0.034977 *
RHU1.R:jet.position.effect                       3.272e-03  2.391e-03  1.368 0.171332
LatLeg1:LongLeg1:S1                             -1.728e+00  1.883e+00 -0.918 0.358662
LatLeg1:LongLeg1:C1                              5.768e+00  2.009e+00  2.871 0.004101 **
LatLeg1:LongLeg1:S2                              1.208e+00  1.749e+00  0.691 0.489848
LatLeg1:LongLeg1:C2                             -2.615e+00  1.617e+00 -1.616 0.106008
LatLeg1:LongLeg1:AltLeg1                        -3.757e+01  7.958e+00 -4.721 2.35e-06 ***
LatLeg1:LongLeg1:Loglag1                        -3.476e-01  3.104e-01 -1.120 0.262859
```

```
LatLeg1:AltLeg1:S1                 9.270e-02  2.486e-01  0.373 0.709231
LatLeg1:AltLeg1:C1                -1.721e+00  3.125e-01 -5.508 3.66e-08 ***
LatLeg1:AltLeg1:S2                -6.053e-02  2.271e-01 -0.266 0.789859
LatLeg1:AltLeg1:C2                 9.064e-01  1.977e-01  4.584 4.58e-06 ***
LongLeg1:AltLeg1:S1                6.305e-02  5.307e-01  0.119 0.905427
LongLeg1:AltLeg1:C1                2.324e+00  6.283e-01  3.699 0.000217 ***
LongLeg1:AltLeg1:S2               -3.829e-01  4.835e-01 -0.792 0.428491
LongLeg1:AltLeg1:C2               -1.450e+00  4.316e-01 -3.359 0.000783 ***
AltLeg1:S1:LatLeg2                 3.697e-01  2.035e-01  1.816 0.069315 .
AltLeg1:C1:LatLeg2                -1.008e+00  2.542e-01 -3.964 7.39e-05 ***
AltLeg1:S2:LatLeg2                 9.477e-03  2.004e-01  0.047 0.962286
AltLeg1:C2:LatLeg2                 4.902e-01  1.857e-01  2.639 0.008313 **
AltLeg1:S1:LongLeg2               -3.107e-02  5.423e-01 -0.057 0.954320
AltLeg1:C1:LongLeg2                2.405e+00  6.108e-01  3.937 8.26e-05 ***
AltLeg1:S2:LongLeg2               -2.055e-01  4.740e-01 -0.434 0.664549
AltLeg1:C2:LongLeg2               -1.345e+00  4.271e-01 -3.148 0.001643 **
LongLeg1:S1:LatLeg2                8.317e+00  1.373e+00  6.058 1.39e-09 ***
LongLeg1:C1:LatLeg2                2.381e+00  1.459e+00  1.633 0.102530
LongLeg1:S2:LatLeg2               -5.118e-01  1.080e+00 -0.474 0.635621
LongLeg1:C2:LatLeg2                3.923e+00  8.726e-01  4.496 6.94e-06 ***
LongLeg1:Loglag1:LatLeg2          -5.582e-01  2.687e-01 -2.078 0.037745 *
LongLeg1:AltLeg1:LatLeg2           3.577e+01  1.094e+01  3.268 0.001084 **
LatLeg1:S1:LongLeg2               -3.164e+00  1.448e+00 -2.185 0.028874 *
LatLeg1:C1:LongLeg2                1.986e+00  1.432e+00  1.387 0.165376
LatLeg1:S2:LongLeg2               -7.750e-01  1.221e+00 -0.635 0.525499
LatLeg1:C2:LongLeg2               -1.817e+00  1.165e+00 -1.560 0.118855
LatLeg1:Loglag1:LongLeg2          -2.745e-01  3.077e-01 -0.892 0.372271
LatLeg1:AltLeg1:LongLeg2          -6.842e+01  1.201e+01 -5.696 1.24e-08 ***
LatLeg1:LongLeg1:AltLeg2          -2.356e+01  6.766e+00 -3.482 0.000498 ***
LongLeg1:LatLeg2:AltLeg2          -2.449e+01  7.948e+00 -3.082 0.002059 **
LatLeg1:LongLeg2:AltLeg2           2.127e+01  6.128e+00  3.470 0.000520 ***
LatLeg1:LongLeg1:AltLeg3           5.097e-01  4.643e+00  0.110 0.912600
LongLeg1:LatLeg2:AltLeg3           8.035e+00  2.855e+00  2.815 0.004886 **
LatLeg1:LongLeg2:AltLeg3          -9.211e+00  3.866e+00 -2.383 0.017184 *
LongLeg1:S1:LatLeg3                2.278e+00  6.487e-01  3.512 0.000445 ***
LongLeg1:C1:LatLeg3                4.898e+00  7.161e-01  6.839 8.11e-12 ***
LongLeg1:S2:LatLeg3                3.173e-02  5.603e-01  0.057 0.954841
LongLeg1:C2:LatLeg3               -7.166e-01  5.458e-01 -1.313 0.189243
LongLeg1:AltLeg1:LatLeg3           1.091e+00  3.508e+00  0.311 0.755742
LatLeg1:S1:LongLeg3               -5.327e+00  9.912e-01 -5.375 7.72e-08 ***
LatLeg1:C1:LongLeg3                1.591e+00  1.079e+00  1.475 0.140351
LatLeg1:S2:LongLeg3                1.079e+00  8.329e-01  1.296 0.195013
LatLeg1:C2:LongLeg3               -2.519e+00  7.702e-01 -3.270 0.001077 **
LatLeg1:AltLeg1:LongLeg3          -3.688e+01  6.172e+00 -5.975 2.32e-09 ***
S1:LatLeg2:LongLeg2                2.405e+00  1.198e+00  2.007 0.044787 *
C1:LatLeg2:LongLeg2               -2.220e+00  1.486e+00 -1.494 0.135133
```

```
S2:LatLeg2:LongLeg2                                    -1.747e+00  1.128e+00  -1.549 0.121345
C2:LatLeg2:LongLeg2                                     4.420e+00  9.556e-01   4.625 3.76e-06 ***
AltLeg1:LatLeg2:LongLeg2                               -1.281e+01  5.030e+00 -2.547 0.010885 *
LatLeg1:LongLeg1:S3                                     4.577e-02  1.679e-01   0.273 0.785166
LatLeg1:LongLeg1:C3                                     2.287e-01  1.397e-01   1.638 0.101516
LatLeg1:LongLeg1:RHU1.R                                -4.240e-02  1.723e-02 -2.460 0.013881 *
LatLeg1:LongLeg1:nino3.4                               -2.684e-01  1.337e-01 -2.008 0.044665 *
LatLeg1:LongLeg1:sht.stdanom                            5.012e-02  9.452e-02   0.530 0.595952
LatLeg1:LongLeg1:SAM                                   -1.451e-01  5.150e-02 -2.818 0.004841 **
LatLeg1:LongLeg1:swio                                  -4.937e-02  8.742e-02 -0.565 0.572240
S1:jet.intensity.stdanom:jet.position.effect            1.016e-01  3.400e-02   2.988 0.002809 **
C1:jet.intensity.stdanom:jet.position.effect            5.185e-02  6.435e-02   0.806 0.420390
SAM:jet.intensity.stdanom:jet.position.effect          -4.960e-03  3.650e-03 -1.359 0.174207
sht.stdanom:jet.intensity.stdanom:jet.position.effect -1.075e-03  9.901e-03 -0.109 0.913511
RHU1.R:jet.intensity.stdanom:jet.position.effect        3.322e-03  2.535e-03   1.311 0.189988
LatLeg1:LongLeg1:AltLeg1:S1                             8.407e-01  4.003e-01   2.100 0.035706 *
LatLeg1:LongLeg1:AltLeg1:C1                             1.528e-02  4.839e-01   0.032 0.974812
LatLeg1:LongLeg1:AltLeg1:S2                             4.197e-01  3.631e-01   1.156 0.247706
LatLeg1:LongLeg1:AltLeg1:C2                            -1.749e-01  3.486e-01 -0.502 0.615850
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
```

Bibliography

Abramowitz, M. and I. Stegun, 1965: *Handbook of mathematical functions: with formulas, graphs and mathematical tables*. Dover, New York.

Akpalu, W. and C. Hassan, R. M. Ringler, 2008: Climate variability and maize yield in South Africa. IFPRI research brief, `http://www.ifpri.org/sites/default/files/publications/rb15_10.pdf`, International Food Policy Research Institute (IFPRI).

Allan, R. P. and B. J. Soden, 2008: Atmospheric warming and the amplification of precipitation extremes. *Science*, **321**, 1481–1484.

Alston, J. M., T. J. Wyatt, P. G. Pardey, M. C. Marra, and C. Chan-Kang, 2000: A meta-analysis of rates of return to agricultural R & D: ex pede Herculem? Research reports 113, `http://www.ifpri.org/sites/default/files/publications/rr113.pdf`, International Food Policy Research Institute (IFPRI).

Ati, O. F., C. J. Stigter, and E. O. Oladipo, 2002: A comparison of methods to determine the onset of the growing season in northern Nigeria. *Int. J. Climatol.*, **22**, 731–742.

Bates, B. C., Z. W. Kundzewicz, S. Wu, and J. P. Palutikof, 2008: Climate change and water. technical paper of the Intergovernmental Panel on Climate Change. Tech. rep., `http://www.ipcc.ch/pdf/technical-papers/climate-change-water-en.pdf`, IPCC Secretariat, Geneva.

Beaulieu, C., T. B. M. J. Ouarda, and O. Seidou, 2007: Synthèse des techniques d'homogénéisation des séries climatiques et analyse d'applicabilité aux séries de précipitations. *Hydro. Sc. J.*, **52**, 18–37.

Berkouwer, S., 2010: Modelling rainfall in South Africa. Using generalized linear models to determine the effects of local and continental covariates on local rainfall patterns, b.Sc. Thesis.

Biasutti, M., A. H. Sobel, and Y. Kushnir, 2006: AGCM precipitation biases in the tropical Atlantic. *J. Clim.*, **19**, 935–958.

Bowman, A. and A. Azzalini, 1997: *Applied smoothing techniques for data analysis: the kernel approach with S–Plus illustrations*. Oxford University Press, Oxford.

Bowman, A. W., M. Giannitrapani, and M. E. Scott, 2009: Spatiotemporal smoothing and sulphur dioxide trends over europe. *J. Royal Stat. Soc. C*, **58**, 737–752.

Brissette, F. P., M. Khalili, and R. Leconte, 2007: Efficient stochastic generation of multi-site synthetic precipitation data. *Journal of Hydrology*, **345 (3-4)**, 121–133.

Brohan, P., J. J. Kennedy, I. Harris, S. F. B. Tett, and P. D. Jones, 2006: Uncertainty estimates in regional and global observed temperature changes: A new data set from 1850. *J. Geophys. Res.*, **111**, doi:10.1029/2005JD006548.

Buishand, T. A., M. V. Shabalova, and T. Brandsma, 2004: On the choice of the temporal aggregation level for statistical downscaling of precipitation. *J. Clim.*, **17**, 1816–1827.

Byerlee, D. and C. K. Eicher, 1997: *Africa's emerging maize revolution*. Lynne Rienner Publishers, Boulder, Colorado.

Chandler, R. E., 2002: GLIMCLIM: Generalized linear modelling for daily climate time series - user guide. Tech. rep., `http://www.ucl.ac.uk/~ucakarc/work/rain_glm.html`, Department of Statistical Science, University College London.

Chandler, R. E., 2005: On the use of generalized linear models for interpreting climate variability. *Environmetrics*, **16**, 699–715.

Chandler, R. E. and S. Bate, 2007: Inference for clustered data using the independence loglikelihood. *Biometrika*, **94**, 167–183.

Chandler, R. E. and E. M. Scott, 2011: *Statistical Methods for Trend Detection and Analysis in the Environmental Sciences*. John Wiley and Sons, Chichester.

Chandler, R. E. and H. S. Wheater, 2002: Analysis of rainfall variability using generalized linear models: A case study from the West of Ireland. *Water Resourc. Res.*, **38 (10)**.

Christensen, J. H., et al., 2007: *2007: Regional Climate Projections*. Cambridge University Press, Cambridge.

Coe, R. and R. D. Stern, 1982: Fitting models to daily rainfall. *J. Appl. Meteorol.*, **21**, 1024–1031.

Collins, M., 2007: Ensembles and probabilities: a new era in the prediction of climate change. *Phil. Trans. R. Soc. A*, **365**, 1957–1970.

Conway, D., A. Persechino, S. Ardoin-Bardin, H. Hamandawana, C. Dieulin, and G. Mahé, 2009: Rainfall and water resources variability in sub-saharan Africa during the twentieth century. *J. Hydrometeor.*, **20**, 41–59.

Cook, C., C. J. C. Reason, and B. C. Hewitson, 2004: Wet and dry spells within particularly wet and dry summers in the South African summer rainfall region. *Clim. Res.*, **26**, 17–31.

Cressie, N. A. C., 1991: *Statistics for spatial data*. J. Wiley, New York.

CRU, 2008: Climate Research Unit Datasets. British Atmospheric Data Centre. `http://badc.nerc.ac.uk/data/cru`.

Dawid, A. P., 1984: Statistical theory: The prequential approach. *J. Royal Stat. Soc., A*, **147**, 278–292.

Dessai, S., X. Lu, and M. Hulme, 2005: Limited sensitivity analysis of regional climate change probabilities for the 21^{th} century. *J. Geoph. Res.*, **110**, doi:10.1029/2005JD005919.

Dilley, M., 2000: Reducing vulnerability to climate variability in southern Africa: The growing role of climate information. *Clim. Change*, **45**, 63–73.

Dimes, J. and P. du Toit, 2008: Quantifying water productivity in rainfed cropping systems in Limpopo province, RSA. Second International Forum on Water and Food, `http://www.waternetonline.ihe.nl/challengeprogram/C13%20Dimes%20WP.pdf`.

Draggan, S., 2010: *Statistical downscaling approach and downscaling of AOGCM climate change projections*. Cutler J. Cleveland, Washington, D.C.: Environmental Information Coalition, National Council for Science and the Environment.

Fauchereau, N., S. Trzaska, M. Rouault, and Y. Richard, 2003: Rainfall variability and changes in southern Africa during the 20^{th} century in the global warming context. *Nat. Haz.*, **29**, 139–154.

Fealy, R. and J. Sweeney, 2007: Statistical downscaling of precipitation for a selection of sites in Ireland employing a generalised linear modelling approach. *Int. J. Climatol.*, **27**, 2083–2094.

Findell, K. L. and T. L. Delworth, 2010: Impact of common sea surface temperature anomalies on global drought and pluvial frequency. *J. Clim.*, **23**, 485–503.

Foley, J. A., S. Levis, I. C. Prentice, D. Pollard, and S. L. Thompson, 1998: Coupling dynamic models of climate and vegetation. *Global Change Biology*, **4 (5)**, 561–579.

Fowler, H. J., S. Blenkinsop, and C. Tebaldi, 2007: Linking climate change modelling to impact studies: recent advances in downscaling techniques for hydrological modelling. *Int. J. Climatol.*, **27**, 1547–1578.

Fox, J., 2002: *An* R *and* S-PLUS *Companion to Applied Regression*. Sage Publications, Thousand Oaks.

Frost, A. J., R. E. Chandler, and M. Segond, 2006: Multi-site downscaling for the Blackwater catchment. Tech. rep., `http://www.ucl.ac.uk/statistics/research/Rainfall/FD2113_rpt4.pdf`, UCL, London.

Furrer, E. M. and R. W. Katz, 2007: Generalized linear modelling approach to stochastic weather generators. *Clim. Res.*, **34**, 129–144.

Gasse, F., F. Chaliè, A. Vincens, M. A. J. Williams, and D. Williamson, 2008: Climatic patterns in equatorial and southern Africa from 30,000 to 10,000 years ago reconstructed from terrestrial and near-shore proxy data. *Quat. Sci. Rev.*, **27**, 2316–2340.

Giannini, A., M. Biasutti, I. M. Held, and A. H. Sobel, 2008: A global perspective on African climate. *Clim. Change*, **90**, 359–383.

Gillett, N. P., T. D. Kell, and P. D. Jones, 2006: Regional climate impacts of the Southern Annular Mode. *Geophys. Res. Lett.*, doi:10.1029/2006GL027721.

Giorgi, F. and R. Francisco, 2000: Evaluating uncertainties in the prediction of regional climate change. *Geophys. Res. Lett.*, **27**, 1295–1298.

Hachigonta, S., C. J. C. Reason, and M. Tadross, 2008: An analysis of onset date and rainy season duration over Zambia. *Theor. Appl. Climatol.*, **91**, 229–243.

Harrison, M. S. J., 1984: A generalized classification of South African summer rain-bearing synoptic systems. *J. Clim.*, **4**, 547–560.

Hastie, T., 2011: *gam: Generalized Additive Models*. URL `http://CRAN.R-project.org/package=gam`, r package version 1.04.

Hawkins, E. and R. Sutton, 2009: The potential to narrow uncertainty in regional climate predictions. *BAMS*, **90**, 1095–1107.

Hewitson, B. C. and R. G. Crane, 1996: Climate downscaling: techniques and application. *Clim. Res.*, **7**, 85–95.

Hoerling, M. P., J. W. Hurrell, and J. Eischeid, 2006: Detection and attribution of twentieth-century northern and southern African rainfall change. *J. Clim.*, **19**, 3989–4008.

Houghton, J. T., Y. Ding, D. J. Griggs, M. Noguer, P. J. van der Linden, X. Dai, K. Maskell, and C. A. Johnson, 2001: *Climate Change 2001: The Scientific Basis. Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Houghton, J. T., L. G. Meira Filho, B. A. Callander, N. Harris, A. Kattenberg, and K. Maskell, 1995: *Climate Change 1995: The Science of Climate Change. Contribution of Working Group I to the Second Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Hulme, M., 1992: Rainfall changes in Africa: 1931-1960 to 1961-1990. *Int. J. Climatol.*, **12**, 685–699.

Hulme, M., D. Conway, P. D. Jones, T. Jiang, E. M. Barrow, and C. Turney, 1995: Construction of a 1961–1990 european climatology for climate change modelling and impact applications. *Int. J. Climatol.*, **15**, 1333–1363.

Hulme, M., R. Doherty, T. Ngara, M. New, and D. Lister, 2001: African climate change: 1990–2100. *Clim. Res.*, **17**, 145–168.

Huntington, T. G., 2006: Evidence for intensification of the global water cycle: Review and synthesis. *J. Hydrol.*, **319**, 83–95.

Husak, G. J., J. Michaelsen, and C. Funk, 2007: Use of the gamma distribution to represent monthly rainfall in Africa for drought monitoring applications. *Int. J. Climatol.*, **27**, 935–944.

IPCC, 2000: Special report on emission scenarios. Tech. rep., `http://www.grida.no/publications/other/ipcc_sr/?src=/climate/ipcc/emission/`, Intergovernmental Panel on Climate Change.

IPCC, 2007: Climate change 2007: Synthesis report. Tech. rep., Intergovernamental Panel on Climate Change.

Jones, P. D., T. Jónsson, and D. Wheeler, 1997: Extension to the North Atlantic Oscillation using early instrumental pressure observations from Gibraltar and south-west Iceland. *Int. J. Climatol.*, **17**, 1433–1450.

Jury, M. R., C. Mc Queen, and K. Levey, 1994: SOI and QBO signals in the African region. *Theor. Appl. Climatol.*, **50**, 103–115.

Jury, M. R., W. B. White, and C. J. C. Reason, 2004: Modelling the dominant climate signals around southern Africa. *Clim. Dynam.*, **23**, 717–726.

Kalnay, E., et al., 1996: The NCEP/NCAR 40-year reanalysis project. *BAMS*, **77**, 437–471.

Kenabatho, P. K., N. R. McIntyre, R. Chandler, and H. Wheater, 2011: Stochastic simulation of rainfall in the semi-arid limpopo basin, botswana. *Int. J. Climat.*, doi: doi.10.1002/joc.2323.

Kigobe, M., N. McIntyre, H. Wheater, and R. Chandler, 2011: Multi-site stochastic modelling of daily rainfall in uganda. *Hydrol. Sci. J.*, **56**, 17–33.

Kistler, R., et al., 2001: The NCEP–NCAR 50-year reanalysis: Monthly means CD-ROM and documentation. *BAMS*, **82**, 247–268.

Kniveton, D. R., R. Layberry, C. J. R. Williams, and M. Peck, 2009: Trends in the start of the wet season over Africa. *Int. J. Climatol.*, **29**, 1216–1225.

Lauer, J. G., P. R. Carter, T. M. Wood, G. Diezel, D. W. Wiersma, R. E. Rand, and M. J. Mlynarek, 1999: Corn hybrid response to planting date in the northern corn belt. *Agron. J.*, **91**, 834–839.

Leith, N. A., 2005a: Forecasting local scale precipitation under scenarios of global climate change: A literature review. *Project FD2113 report n.1.*

Leith, N. A., 2005b: Using generalised linear models to simulate daily rainfall under scenarios of climate change. *Project FD2113 report n.2.*

Levey, K. M. and M. R. Jury, 1996: Composite intraseasonal oscillations of convection over southern Africa. *J. Clim.*, **9**, 1910–1920.

Lin, J., 2007: The double-ITCZ problem in IPCC AR4 coupled GCMs: Ocean-atmosphere feedback analysis. *J. Clim.*, **20**, 4497–4525.

Lobell, D. B., B. B. Marshall, C. Tebaldi, M. D. Mastrandrea, W. P. Falcon, and R. L. Naylor, 2008: Prioritizing climate change adaptation needs for food security in 2030. *Science*, **319**, 607–610.

Lyon, B. and S. J. Mason, 2007: The 1997–98 summer rainfall season in southern Africa. Part I: observations. *J.Clim.*, **20**, 5134–5148.

Lyon, B. and S. J. Mason, 2009: The 1997–98 summer rainfall season in southern Africa. Part II: model simulations and coupled model forecasts. *J.Clim.*, **29**, 3802–3818.

Manatsa, D., W. Chingombe, and C. H. Matarira, 2008: The impact of the positive indian ocean dipole on the Zimbabwe drought. *Int. J. Climatol.*, **28**, 2011–2029.

Maraun, D., et al., 2010: Precipitation downscaling under climate change recent developments to bridge the gap between dynamical models and the end users. *Rev. Geoph.*, **48**, doi:10.1029/2009RG000314.

Marshall, G. J., 2003: Trends in the Southern Annular Mode from observations and reanalysis. *J. Clim.*, **16**, 4134–4143.

Mason, S. J., 1995: Sea-surface temperature – South African rainfall associations. *Int. J. Climatol.*, **15**, 119–135.

Mason, S. J., 2001: El-Niño, climate change and southern African climate. *Environmetrics*, **12**, 327–345.

Mason, S. J. and M. R. Jury, 1997: Climatic variability and change over southern Africa: a reflection on underlying processes. *Progr. Phys. Geog.*, **21**, 23–50.

Mason, S. J. and P. D. Tyson, 2000: *The occurrence and predictability of droughts over Southern Africa*, Vol. I. Routledge.

Matarira, C. H., 1990: Drought over Zimbabwe in a regional and global context. *Int. J. Climatol.*, **10**, 609–625.

McCabe, G. J. and M. A. Palecki, 2006: Multidecadal climate variability of global lands and oceans. *J. Climatol.*, **26**, 849–865.

McHugh, M. J. and J. C. Rogers, 2001: North Atlantic Oscillation influence on precipitation variability around the southeast African convergence zone. *J. Clim.*, **14**, 3631–3642.

Meehl, G. A., et al., 2007: *Global Climate Projections*. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Mitchell, T. D. and P. D. Jones, 2005: An improved method of constructing a database of monthly climate observations and associated high-resolution grids. *Int. J. Climatol.*, **25**, 693–712.

Montgomery, D. C. and E. Peck, 1992: *Introduction to linear regression analysis*. Wiley, New York.

Morgan, M. G., L. F. Pitelka, and E. Shevliakova, 2001: Elicitation of expert judgements of climate change impacts on forest ecosystems. *Clim. Ch.*, **49**, 279–307.

Morishima, W. and I. Akasaka, 2010: Seasonal trends of rainfall and surface temperature over southern africa. *Afric. Study Mon.*, **40**, 67–76.

Nicholson, S. E. and J. Kim, 1997: The relationship of ENSO to African rainfall. *Int. J. Climatol.*, **17**, 117–135.

Northrop, P., 2010: Using statistics to assess climate uncertainty, oral presentation at 11$^{\text{th}}$ IMSC.

Parry, M. L., O. F. Canziani, and J. P. Palutikof, 2007: *Technical Summary*. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Pohl, B., Y. Richard, and N. Fauchereau, 2007: Influence of the Madden–Julian Oscillation on southern African summer rainfall. *J. Clim.*, **20**, 4227–4242.

Quinn, G. P. and M. J. Keough, 2002: *Experimental design and data analysis for biologists*. Cambridge University Press, Cambridge.

R Development Core Team, 2006: *R: A Language and Environment for Statistical Computing*. Vienna, Austria, R Foundation for Statistical Computing, ISBN 3-900051-07-0.

Raes, D., A. Sithole, A. Makarau, and J. Milford, 2004: Evaluation of first planting dates recommended by criteria currently used in Zimbabwe. *Agric. and For. Meteor.*, **125**, 177–185.

Räisänen, J. and T. N. Palmer, 2001: A probability and decision model analysis of a multimodel ensemble of climate change simulations. *J. Clim.*, **14**, 3212–3226.

Rayner, N. A., D. E. Parker, E. B. Horton, C. K. Folland, L. V. Alexander, D. P. Rowell, E. C. Kent, and A. Kaplan, 2003: Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *Geophys. Res. Lett.*, **108**, doi:10.1029/2002JD002670.

Reason, C. J. C., 2001: Subtropical Indian Ocean SST dipole events and southern African rainfall. *Geophys. Res. Lett.*, **28**, 2225–2227.

Reason, C. J. C., S. Hachigonta, and R. F. Phaladi, 2005: Interannual variability in rainy season characteristics over the Limpopo region of southern Africa. *Int. J. Climatol.*, **25**, 1835–1853.

Reason, C. J. C. and D. Jagadheesha, 2005a: A model investigation of recent ENSO impacts over southern Africa. *Meteorol. Atmos. Phys.*, **89**, 181–205.

Reason, C. J. C. and D. Jagadheesha, 2005b: Relationship between south Atlantic SST variability and atmospheric circulation over the South African region during austral winter. *J. Clim.*, **19**, 3059–3075.

Reason, C. J. C., W. Landman, and W. Tennant, 2006: Seasonal to decadal prediction of southern African climate and its links with variability of the Atlantic Ocean. *BAMS*, **87**, 941–955.

Reason, C. J. C., W. A. Landman, M. Tadross, W. Tennant, and M. J. Kgatuke, 2004: Seasonal to decadal predictability and prediction of southern African climate. *CLIVAR exchange*, **9**, 21–23.

Reason, C. J. C. and M. Rouault, 2002: ENSO–like decadal variability and South African rainfall. *Geophys. Res. Lett.*, **29**, doi:10.1029/2002GL014663.

Reason, C. J. C. and M. Rouault, 2005: Links between the Antarctic Oscillation and winter rainfall over western South Africa. *Geophys. Res. Lett.*, **32**, doi: 10.1029/2005GL022419.

Richard, Y., N. Fauchereau, I. Poccard, M. Rouault, and S. Trzaska, 2001: 20^{th} century droughts in southern Africa - spatial and temporal variability, teleconnections with oceanic and atmospheric conditions. *Int. J. Climatol.*, **21**, 873–895.

Richard, Y., S. Trzaska, P. Roucou, and M. Rouault, 2000: Modification of the southern African rainfall variability/El Niño Southern Oscillation relationship. *Clim. Dyn.*, **16**, 883–895.

Rocha, A. and I. Simmonds, 1997a: Interannual variability of south-eastern African summer rainfall. Part I: Relationships with air-sea interaction processes. *Int. J. Climatol.*, **17**, 235–265.

Rocha, A. and I. Simmonds, 1997b: Interannual variability of south-eastern African summer rainfall. Part II: Modelling the impact of sea-surface temperatures on rainfall circulation. *Int. J. Climatol.*, **17**, 267–290.

Rogers, J. C. and H. van Loon, 1982: Spatial variability of sea level pressure and 500-mb height anomalies over the Southern Hemisphere. *Mon. Weath. Rev.*, **110**, 1375–1392.

Ropelewski, C. F. and P. D. Jones, 1987: An extension of the Tahiti-Darwin Southern Oscillation Index. *Mon. Weath. Rev.*, **115**, 2161–2165.

Roy, A. and D. Clarke, 1988: *Astronomy principles and practice.* Adam Hilger, Bristol.

Rummukainen, M., 1997: Methods of statistical downscaling of GCM simulations. Reports Meteorology and Climatology 80, Swedish Meteorological and Hydrological Institute, Sweden.

Saji, N. H., B. N. Goswami, P. N. Vinayachandran, and T. Yamagata, 1999: A dipole mode in the tropical Indian Ocean. *Nature*, **401**, 360–363.

Schubert, S., et al., 2009: A U.S. CLIVAR project to assess and compare the responses of global climate models to drought-related SST forcing patterns: Overview and results. *J. Clim.*, **22**, 5251–5272.

Shongwe, M. E., G. J. van Oldenborgh, B. J. J. M. van den Hurk, B. de Boer, and C. A. S. Coelho, 2009: Projected changes in mean and extreme precipitation in Africa under global warming. Part I: Southern Africa. *J. Clim.*, **22**, 3819–3837.

Shukla, J. and Y. Mintz, 1982: Influence of land-surface evapotranspiration on the earth's climate. *Science*, **215**, 1498–1501.

Slingo, J. M., A. J. Challinor, B. J. Hoskins, and T. R. Wheeler, 2005: Introduction: food crops in a changing climate. *Philosop. Trans. Royal Soc., Series B*, **360**, 1983–1989.

Smale, M. and T. S. Jayne, 2003: Maize in eastern and southern Africa:'Seeds' of success in retrospect. EPTD discussion paper no.97, Environment and Production Technology Division IFPRI, Washington, D.C., U.S.A.

Smith, T. M., R. W. Reynolds, T. C. Peterson, and J. Lawrimore, 2008: Improvements to NOAA's historical merged land-ocean surface temperature analysis (1880–2006). *J. Clim.*, **21**, 2283–2296.

Solomon, S., D. Qin, M. Manning, Z. Chen, M. Marquis, K. B. Averyt, M. Tignor, and H. L. Miller, 2007: *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press.

Srikanthan, R. and T. A. McMahon, 2001: Stochastic generation of annual, monthly and daily climate data: a review. *Hydrol. Earth Syst. Sc.*, **5**, 653–670.

Stainforth, D. A., M. R. Allen, E. R. Tredger, and L. A. Smith, 2007: Confidence, uncertainty and decision-support relevance in climate predictions. *Phil. Trans. R. Soc. A*, **365**, 2145–2161.

Ster, D. I. and R. K. Kaufmann, 2000: Detecting a global warming signal in hemispheric temperature series: a structural time series analysis. *Clim. Change*, **47**, 411–438.

Sterl, A., 2004: Notes and correspondence on the (in)homogeneity of reanalysis products. *J. Clim.*, **17**, 3866–3873.

Stern, R. D., M. D. Dennett, and I. C. Dale, 1982: Analysing daily rainfall measurements to give agronomically useful results. I. direct methods. *Experim. Agric.*, **18**, 223–236.

Tadross, M., et al., 2007: Changes in growing-season rainfall characteristics and down-scaled scenarios of change over southern Africa: implications for growing maize. Tech. rep., `http://www.wikiadapt.org/filestore/wikiADAPT/Growing_Season_Changes_Southern_Africa.pdf`, IPCC regional Expert Meeting on Regional Impacts, Adaptation, Vulnerability, and Mitigation.

Tadross, M., et al., 2009: Growing-season rainfall and scenarios of future change in southeast Africa: implications for cultivating maize. *Clim. Res.*, **40**, 147–161.

Tebaldi, C. and R. Knutti, 2007: The use of the multi-model ensemble in probabilistic climate projections. *Phil. Trans. R. Soc. A*, **365**, 2053–2075.

Tebaldi, C., R. L. Smith, D. Nychka, and L. O. Mearns, 2005: Quantifying uncertainty in projections of regional climate change: A bayesian approach to the analysis of multi-model ensembles. *J. Clim.*, **18**, 1524–1540.

Tennant, W. J., 2004: Considerations when using pre-1979 NCEP/NCAR reanalyses in the southern hemisphere. *Geophys. Res. Lett.*, **31**, doi:10.1029/2004GL019751.

Tennant, W. J. and B. C. Hewitson, 2002: Intra-seasonal rainfall characteristics and their importance to the seasonal prediction problem. *Int.J. Climat.*, **22**, 1033–1048, doi:10.1002/joc.778.

Todd, M. C. and R. Washington, 1999: Circulation anomalies associated with tropical-temperate troughs in southern Africa and the south west Indian Ocean. *Clim. Dyn.*, **15**, 937–951.

Tyson, P. D., 1981: Atmospheric circulation variations and the occurrence of extended wet and dry spells over southern Africa. *J. Climatol.*, **1**, 115–130.

Tyson, P. D. and R. A. Preston-White, 2000: *The weather and climate of Southern Africa*. Oxford University Press, Oxford.

Van Heerden, J. and J. J. Taljaard, 1998: *Africa and surrounding waters in Meteorology of the Southern Hemisphere*, Meteorological Monographs, Vol. 27. American Meteorological Society.

Van Oldenborgh, G. J., L. A. te Raa, H. A. Dijkstra, and S. Y. Philip, 2009: Frequency- or amplitude-dependent effects of the Atlantic meridional overturning on the tropical Pacific Ocean. *Ocean Sci.*, **5**, 293–301.

Vicente-Serrano, S. M., S. Beguería, J. I. López-Moreno, M. A. García-Vera, and P. Stepanek, 2010: A complete daily precipitation database for northeast Spain: reconstruction, quality control, and homogeneity. *Int. J. Climatol.*, **30**, 1146–1163.

Vigaud, N., Y. Richard, M. Rouault, and N. Fauchereau, 2008: Moisture transport between the South Atlantic Ocean and southern Africa: relationships with summer rainfall and associated dynamics. *Clim. Dyn.*, **5**, doi:10.1007/s00382-008-0377-7.

Vincent, L. A., et al., 2011: Observed trends in indices of daily and extreme temperature and precipitation for the countries of the western indian ocean, 1961–2008. *J. Geophys. Res.*, **116**, doi:10.1029/2010JD015303.

Viney, N. R., et al., 2005: Ensemble modelling of the hydrological impacts of land use change. *MODSIM 2005 International Congress on Modelling and Simulation*, A. Zerger and R. Argent, Eds., 2967–2973.

Walker, N. D., 1990: Links between south African summer rainfall and temperature variability of the Agulhas and Benguela current systems. *J. Geoph. Res.*, **95**, 3297–3319.

Ward, M. N., 1998: Diagnosis and short-lead time prediction of summer rainfall in tropical north Africa at interannual and multidecadal timescales. *J. Clim.*, **11**, 3167–3191.

Washington, R. and A. Preston, 2006: Extreme wet years over southern Africa: Role of Indian Ocean sea surface temperatures. *J. Geoph. Res.*, **111**, doi:10.1029/2005JD006724.

Washington, R. and M. C. Todd, 1999: Circulation anomalies associated with tropical-temperate troughs over southern Africa and the southwest Indian Ocean. *Clim. Dyn.*, **15**, 937–951.

Washington, R., et al., 2006: African climate change, taking the shorter route. *BAMS*, **87**, 1355–1366.

Weaver, A. J. and F. W. Zwiers, 2000: Uncertainty in climate change. *Nature*, **407**, 571–572.

Wilby, R. L., S. P. Charles, E. Zorita, B. Timbal, P. Whetton, and L. O. Mearns, 2004: Guidelines for use of climate scenarios developed from statistical downscaling methods. Tech. rep., `http://www.ipcc-data.org/guidelines/dgm_no2_v1_09_2004.pdf`, Intergovernmental Panel on Climate Change. Supporting Material.

Wilby, R. L. and I. Harris, 2006: A framework for assessing uncertainties in climate change impacts: Low-flow scenarios for the River Thames, UK. *Water Resour. Res.*, **42**, doi:10.1029/2005WR004065.

Wilby, R. L. and T. M. L. Wigley, 1997: Downscaling general circulation model output: a review of methods and limitations. *Progr. in Phys. Geog.*, **21**, 530–548.

Wilby, R. L. and T. M. L. Wigley, 2000: Precipitation predictors for downscaling: observed and general circulation model relationships. *Int. J. Climatol*, **20**, 641–661.

Wilks, D. S. and R. L. Wilby, 1999: The weather generation game: a review of stochastic weather models. *Progr. in Phys. Geog.*, **23**, 329–357.

Williams, C. J. R., D. R. Kniveton, and R. Layberry, 2008: Influence of south Atlantic sea surface temperatures on rainfall variability and extremes over southern Africa. *J. Clim.*, **21**, 6498–6520.

Williams, P. D., E. Guilyardi, R. Sutton, J. Gregory, and G. Madec, 2007: A new feedback on climate change from the hydrological cycle. *Geophys. Res. Lett.*, **34**, doi: 10.1029/2007GL029275.

Yan, Z., S. Bate, R. Chandler, V. Isham, and H. Wheater, 2002: An analysis of daily maximum wind speed in northwestern Europe using generalized linear models. *J. Clim.*, **15**, 2073–2088.

Yan, Z., S. Bate, R. E. Chandler, V. Isham, and H. Wheater, 2006: Changes in extreme wind speeds in NW Europe simulated by generalized linear models. *Theor. Appl. Climatol.*, **83**, 121–137.

Yang, C., R. E. Chandler, V. Isham, and H. S. Wheater, 2005: Spatial-temporal rainfall simulation using generalized linear models. *Water Resourc. Res.*, **41 (W11415)**, doi: 10.1029/2004WR003739.

Zhang, Y., J. M. Wallace, and D. S. Battisti, 1997: ENSO-like interdecadal variability: 1900–93. *J. Clim.*, **10**, 1004–1020.

Ziervogel, G. and R. Calder, 2003: Climate variability and rural livelihoods: assessing the impact of seasonal climate forecast in Lesotho. *Area*, **35**, 403–417.